Microscopic Image Classification via $\mathbb{C}$WT-Based Covariance Descriptors Using Kullback-Leibler Distance

Furkan Keskin and A. Enis Çetin
Department of Electrical and Electronics Engineering
Bilkent University, Ankara, Turkey
Email: keskin@ee.bilkent.edu.tr, cetin@bilkent.edu.tr

Tulin Ersahin and Rengul Çetin-Atalay
Department of Molecular Biology and Genetics
Bilkent University, Ankara, Turkey
Email: rengul@bilkent.edu.tr, ersahin@fen.bilkent.edu.tr

Abstract—In this paper, we present a novel method for classification of cancer cell line images using complex wavelet-based region covariance matrix descriptors. Microscopic images containing irregular carcinoma cell patterns are represented by randomly selected subwindows which possibly correspond to foreground pixels. For each subwindow, a new region descriptor utilizing the dual-tree complex wavelet transform coefficients as pixel features is computed. $\mathbb{C}$WT as a feature extraction tool is preferred primarily because of its ability to characterize singularities at multiple orientations, which often arise in carcinoma cell lines, and approximate shift invariance property. We propose new dissimilarity measures between covariance matrices based on Kullback-Leibler (KL) divergence and $\mathcal{L}_2$-norm, which turn out to be as successful as the classical KL divergence, but with much less computational complexity. Experimental results demonstrate the effectiveness of the proposed image classification framework. The proposed algorithm outperforms the recently published eigenvalue-based Bayesian classification method.

I. INTRODUCTION

In this article, we propose a method for classification of microscopic cancer cell line images using complex wavelet transform ($\mathbb{C}$WT) based region covariance matrix descriptors and new efficient distance measures for comparing covariance matrices. New distance measures which are based on calculating Kullback-Leibler (KL) distance or $\mathcal{L}_2$-norm using normalized matrix entries turn out to yield not only high classification performances but also have much less computational complexity than commonly used KL distance in this image classification problem.

Automatic classification of cancer cell line images is an important problem because of its potential use in cancer research. Automated morphological analysis of cancer cell lines reveals the characteristics of gene expression profiles and the degree of invasiveness of tumor cells [1]. For now, human cell line recognition is performed by using short tandem repeat (STR) analysis as a standard tool. However, an expert is needed for this task and conducting STR is very time-consuming. Predicting labels of cancer cell lines in a fast and accurate manner via a pattern classification approach will greatly enhance biologists' ability to identify different types of cell lines without the need to scrutinize each and every biomedical image one by one. Sample images from each of five classes of cell lines, which are *BT-20*, *Cama-1*, *Focus*, *HepG2* and *huh7*, used in this study are shown in Figures 1-5.

The dual-tree complex wavelet transform (DT-CWT) has been recently used in various signal and image processing applications [2], [3], [4], [5]. It has desirable properties such as shift invariance, directional selectivity and lack of aliasing. In the dual-tree $\mathbb{C}$WT, two maximally decimated discrete wavelet transforms are executed in parallel, where the wavelet functions of two different trees form an approximate Hilbert transform pair [6]. Low-pass analysis filters in real and imaginary trees must be offset by half-sample in order to have one wavelet basis as the approximate Hilbert transform of the other wavelet basis [7]. Analyticity allows one-dimensional DT-CWT to be approximately shift-invariant and free of aliasing artifacts often encountered in DWT-based processing. Two-dimensional DT-CWT is also directionally selective in six different orientations, namely, $\{\pm 15, \pm 45, \pm 75\}$.

Microscopic cancer cell line images contain significant amount of oriented singularities. Recently, a Bayesian classification method that uses the sparsity in a transform domain is developed to classify cancer cell lines [1]. Attributes like orientation selectivity and shift invariance render DT-CWT a good choice for the processing of microscopic images with lots of edge- or ridge-like singularities. We incorporate the complex wavelet transform into recently proposed region covariance descriptors [8] for feature extraction from microscopic images. In the region covariance framework each pixel is mapped to a set of pixel properties whose variances and correlations with one another are utilized in a covariance matrix as region descriptor. We use DT-CWT complex coefficient magnitudes in detail subbands as pixel features and compute covariance descriptors. Augmenting covariance matrices with directional information through the use of 2-D DT-CWT helps to improve the discriminative power of descriptors. Each microscopic image is represented by covariance matrices of certain number of subwindows sampled randomly from the whole image. The k-nearest neighbor classifier is employed for image classification where KL divergence and our proposed distance measures are used for calculation of distances between covariance matrices.

This paper is organized as follows. In Section 2, feature extraction step comprising image decomposition method by DT-CWT, random subwindow selection and construction of covariance matrices is described. In Section 3, KL distance and $\mathcal{L}_2$-norm based covariance classification algorithms by utilizing weak k-NN classifiers are explained. Section 4 presents the experimental results and Section 5 concludes the paper.
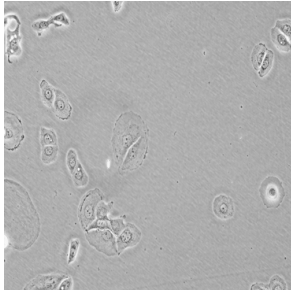
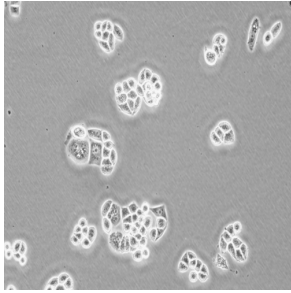Fig. 1. Sample image from *BT-20* class

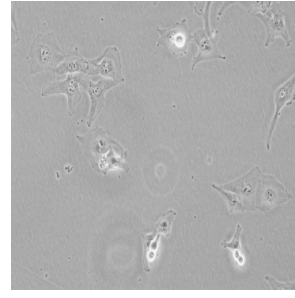Fig. 2. Sample image from *Cama-1* class
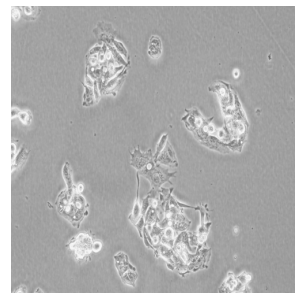

Fig. 3. Sample image from *Focus* class

Fig. 4. Sample image from *HepG2* class

II. Feature Extraction

A. 2-D DT-CWT of Microscopic Images

2-D DT-CWT of an image is obtained by four real separable transforms [9]. Real-part and imaginary-part analysis filters are applied successively to rows and columns of the image. By addition and subtraction of corresponding detail subbands, we obtain a total of 16 subbands consisting of 6 real detail subbands, 6 imaginary detail subbands and 4 approximation subbands. Two-dimensional dual-tree decomposition is an oversampled transform with a redundancy factor of 4 (2^d for d-dimensional signals). In our work, we perform two-level 2-D DT-CWT decomposition of each biomedical image of size mxn and use only the 2nd level detail subband coefficients to better exploit the analyticity of DT-CWT. Each subband at the 2nd level is of size $m/4xn/4$. The original image is lowpass filtered with $[1/4, 1/2, 1/4]$ filters and downsampled by 4 in both directions to obtain a single intensity image $I_a(x, y)$ which represents the original image and will be used as the image to be classified. Let $W_\theta^R(x, y)$ and $W_\theta^{Im}(x, y)$ denote, respectively, the real and imaginary part of the 2nd level complex wavelet coefficient at the position (x,y) corresponding to directional detail subbands at orientation θ, where $\theta \in \{\pm 15, \pm 45, \pm 75\}$. The magnitude of the complex wavelet coefficent is then given by

$$M_\theta(x, y) = \sqrt{W_\theta^R(x, y)^2 + W_\theta^{Im}(x, y)^2} \tag{1}$$

Hence, for each pixel in the average image $I_a(x, y)$, six complex wavelet coefficient magnitudes $M_\theta(x, y)$ representing six different orientations of DT-CWT are extracted. These magnitudes will be utilized as features in the covariance matrix computation for randomly sampled regions of the image $I_a(x, y)$.

B. Complex Wavelet-Based Region Covariance

Successfully employed in pedestrian detection [10] and flame detection [11], covariance descriptors enable the combination of different features over an image region of interest. Given an intensity image I of size mxn, we define a mapping ϕ from image domain to feature domain as

$$F(x, y) = \phi(I, x, y) \tag{2}$$

where each pixel (x,y) is mapped to a set of features and F is the $mxnxd$ dimensional feature function. For a given subwindow R consisting of n pixels, let $(\mathbf{f_k})_{k=1...n}$ be the d-dimensional feature vectors extracted from R. Then, the covariance matrix of region R can be computed as

$$\mathbf{C} = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{f_k} - \mu)(\mathbf{f_k} - \mu)^T \tag{3}$$

where μ is the mean of the feature vectors inside the region R. The covariance matrix is symmetric positive-definite and of size dxd. There exists a very efficient multiplier-less implementation of covariance descriptors, called co-difference matrices, which have been shown to yield comparable performances to the original ones [12].

Since cancer cell line images contain lots of flat, background-like regions, it is not reasonable to compute the covariance matrix over the whole image region. Exclusion of background regions in covariance computation helps to increase the class separability in classification. Segmentation may be applied as a preprocessing step to distinguish between background and foreground regions; however, it does not lead to decent image regions because microscopic images have very distinct characteristics as compared to real-world structured scenes.

We propose to use random subwindow selection method for cancer cell line images as in [13]. Each image $I_a(x, y)$ is represented by possibly overlapping s square subwindows selected at random locations and with random edge lengths. The edge length of the largest possible subwindow is equal to that of the shorter edge of $I_a(x, y)$, while the size of the smallest possible subwindow is 10 times lower than that of the largest one. We enforce a variance
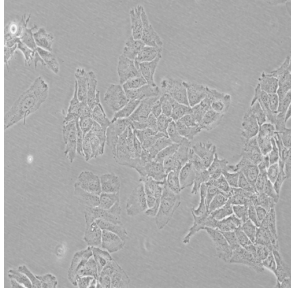
Fig. 5. Sample image from *huh7* class

constraint on the selected subwindows to avoid gathering background regions. A randomly chosen subwindow is discarded if its variance is below an image-dependent threshold, which is determined to be the variance of the whole image. Random window selection process continues until the total number of windows becomes s. We compute a covariance matrix for each subwindow and an image is represented by s covariance matrices. Random sampling approach avoids the need to scan the whole image to regularly take samples, which is computationally expensive.

With

$$\mathbf{M}_\theta(\mathbf{x}, \mathbf{y}) = [M_{\theta_1}(x, y)...M_{\theta_6}(x, y)] \quad (4)$$

where $\theta_1...\theta_6$ corresponds to the six orientations of DT-CWT detail subbands $\{\pm 15, \pm 45, \pm 75\}$, and $M_\theta(x, y)$ is as defined in Equation (1), we employ two different feature mapping functions to compute covariance matrices:

$$\phi_1(I, x, y) = [I_a(x, y)|I_x||I_y||I_{xx}||I_{yy}|]^T, \quad (5)$$

and

$$\phi_2(I, x, y) = [I_a(x, y)|I_x||I_y||I_{xx}||I_{yy}|\mathbf{M}_\theta(\mathbf{x}, \mathbf{y})]^\mathbf{T} \quad (6)$$

where $|I_x|$ and $|I_{xx}|$ denote the first- and second-order derivatives at (x, y) of the image I_a. The sizes of the covariance matrices generated using the mapping functions in (5) and (6) are $5x5$ and $11x11$, respectively. Using derivative features as in (5) is prevalent in region description-related tasks [8], [11]. Incorporating complex wavelet coefficients into image region description in (6) allows us to enhance the discriminative power of covariance matrices, which, in turn, contain directional information at eight different orientations. To the best of our knowledge, no studies exist that employ complex wavelets for describing an image region through covariance matrices.

III. CLASSIFICATION ALGORITHM

The image signature is composed of s covariance matrices of the same size, computed using one of the feature vectors in Equations (5) and (6). We calculate the distance between covariance matrices using both the Kullback-Leibler (KL) divergence and our proposed distance measures. The KL divergence between two covariance matrices $\mathbf{C_1}$ and $\mathbf{C_2}$ is calculated as follows [14]

$$d_{KL}(\mathbf{C_1}, \mathbf{C_2}) = D_{KL}(\mathcal{N}_1||\mathcal{N}_2) \quad (7)$$

$$= \frac{1}{2}\left(log\left(\frac{|\mathbf{C_2}|}{|\mathbf{C_1}|}\right) + tr(\mathbf{C_2}^{-1}\mathbf{C_1}) - d\right) \quad (8)$$

where it is assumed that $\mathbf{C_1}$ and $\mathbf{C_2}$ are the covariance matrices of multivariate Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$ with zero mean

vectors, respectively, and $D_{KL}(p||q)$ is the KL divergence between distributions p and q.

We propose a new distance measure between covariance matrices $\mathbf{C_1}$ and $\mathbf{C_2}$ as follows. First, we take the exponential of all entries of the dxd covariance matrix $\mathbf{C} = (c_{ij})$ to make all entries positive. The resultant covariance matrix $\mathbf{C_e} = (e^{c_{ij}})$ is treated as a bivariate probability density function where each entry denotes the density at the corresponding 2-D histogram bin. Then, the entries of $\mathbf{C_e}$ are normalized to have the sum over all entries equal to 1, leading to a valid density. The normalized matrix is denoted as $\mathcal{C}_e$. The distance between $\mathbf{C_1}$ and $\mathbf{C_2}$ is then calculated as the KL distance between two bivariate distributions with discrete pdfs denoted by $\mathcal{C}_{e_1}$ and $\mathcal{C}_{e_2}$, respectively:

$$d_{KL_e}(\mathbf{C_1}, \mathbf{C_2}) = D_{KL}(\mathcal{C}_{e_1}, \mathcal{C}_{e_2}) = \sum_{i=1}^{d}\sum_{j=1}^{d}\mathcal{C}_{e_1}(i, j)log\left(\frac{\mathcal{C}_{e_1}(i, j)}{\mathcal{C}_{e_2}(i, j)}\right) \quad (9)$$

$\mathcal{L}_2$-norm-based distance can also be used to calculate dissimilarity between $\mathbf{C_1}$ and $\mathbf{C_2}$, or, equivalently, $\mathcal{C}_{e_1}$ and $\mathcal{C}_{e_2}$:

$$d_2(\mathbf{C_1}, \mathbf{C_2}) = D_{\mathcal{L}_2}(\mathcal{C}_{e_1}, \mathcal{C}_{e_2}) = \sum_{i=1}^{d}\sum_{j=1}^{d}(\mathcal{C}_{e_1}(i, j) - \mathcal{C}_{e_2}(i, j))^2 \quad (10)$$

The final distance measure we propose is to use the KL distance directly between the matrix entries by considering them to be densities after normalization as in (9). If there is a negative entry, all entries are increased by an amount slightly greater than the absolute value of the smallest entry before normalization so that matrix represents a valid pdf with no empty bins. With the normalized matrices denoted by $\mathcal{C}_1$ and $\mathcal{C}_2$, the distance is given by

$$d_3(\mathbf{C_1}, \mathbf{C_2}) = D_{KL}(\mathcal{C}_1, \mathcal{C}_2) = \sum_{i=1}^{d}\sum_{j=1}^{d}\mathcal{C}_1(i, j)log\left(\frac{\mathcal{C}_1(i, j)}{\mathcal{C}_2(i, j)}\right) \quad (11)$$

The distance measures (9)-(11) can be calculated efficiently since they only require logarithm and exponential operations, which can be done using look-up tables in an efficient manner. As the covariance matrices are symmetric, we have only $d(d+1)/2$ summands for computation of distances in (9)-(11). These measures do not require computation of matrix inverses and determinants, which are computationally expensive operations, as in the case of Equation (8).

For each test covariance matrix $\mathbf{C}$, the distances from all training covariance matrices are calculated using one of the distance measures in Equations (8)-(11). According to the class labels of the k nearest ones, the posterior probability for each class i ($i = 1...5$) is estimated by $p(i|\mathbf{C}) = \frac{k_i}{k}$ where k_i is the number of votes collected by class i. We mark an image subwindow as unreliable, if $max_i \frac{k_i}{k} \leq \frac{1}{2}$, and discard it. Thus, for each test image we have s - #(discarded windows) weak k-NN classifiers generating posterior probabilities for five different classes. In combining classifiers for a mixture-of-experts model, we employ the sum rule as classifier fusion strategy [15]. Posterior probabilities are summed over the classifiers and the image in question is assigned the label of the class with the largest sum.

IV. EXPERIMENTS AND RESULTS

The dataset used in this study consists of 50 microscopic human carcinoma cell line images with each of the five classes having 10 images. The *BT-20* class denotes a basal breast carcinoma cell line whereas *Cama-1* is a luminal breast carcinoma cell line. *Focus* is a poorly-differentiated hepatocellular carcinoma cell line, and *HepG2*

and *huh7* are well-differentiated hepatocellular carcinoma cell lines. All the images in the dataset were acquired at 20x magnification. The size of each image is 3096x4140 pixels.

To make a fair comparison with [1], we perform experiments by duplicating the test environment of [1]. By adopting the same validation strategy, we take one image as the test image and the remaining ones as the training images for each class. We run ten experiments, choosing each image as the test image only once for each class, and obtain the average image classification accuracy. The number of selected random subwindows is taken to be $s = 100$. In each experiment, there are 45*s = 4500 training covariance matrices and 5*s = 500 test covariance matrices. The training set does not contain any samples from the images used in the testing set. We repeat the above procedure for two different feature mapping functions in Equations (5) and (6) and for three different k values. The results using the distance measures (8)-(11) are shown in Table I.

Distance measure	Feature mapping function	$k = 1$	$k = 5$	$k = 10$	Eigenvalue -based method [1]
(8)	$\phi_1(I, x, y)$	96	96	96	
	$\phi_2(I, x, y)$	100	100	98	
(9)	$\phi_1(I, x, y)$	92	94	94	
	$\phi_2(I, x, y)$	96	98	98	
(10)	$\phi_1(I, x, y)$	92	94	94	90
	$\phi_2(I, x, y)$	98	96	98	
(11)	$\phi_1(I, x, y)$	96	92	92	
	$\phi_2(I, x, y)$	98	94	92	

TABLE I

AVERAGE CLASSIFICATION ACCURACIES (IN %) OF CARCINOMA CELL LINE IMAGES OVER 10 RUNS

It can be seen from Table I that complex wavelet coefficient magnitudes based covariance classification method using $\phi_2(I, x, y)$ outperforms the classical covariance method $\phi_1(I, x, y)$ using only intensity and derivative features. Exploitation of directional information at six different orientations through the use of DT-CWT boosts the image recognition accuracy. Approximate shift invariance property of DT-CWT also adds a certain level of robustness to feature extraction step since it is capable of accurately localizing singularities without causing undesirable positive and negative oscillations around them. With both DT-CWT based and derivative features based mapping functions, our algorithm outperforms the eigenvalue-based Bayesian classification approach proposed in [1], which yielded 90% image classification accuracy.

We demonstrate also that the new dissimilarity measures (9)-(11) compare favorably with the commonly used KL distance (8). Classification accuracies obtained through (9)-(11) are very close to those obtained by using (8). Hence, it transpires that our distance measures can achieve almost the same performance with a much lower computational cost. In addition, as observed from Table I, the distance measures (9)-(11) that we propose are superior to the eigenvalue-based method [1] in that they avoid the computation of eigenvalues of covariance matrices and perform more efficient distance calculation while yielding higher microscopic image classification accuracies.

V. CONCLUSION

In this paper, we demonstrate that automatic classification of microscopic carcinoma cell line images can be reliably performed using complex wavelet-based covariance descriptors. Perfect classification results (100%) are obtained for DT-CWT-based feature vectors, which reveals that the discriminative power of covariance descriptors can be enhanced significantly by incorporating directionally selective DT-CWT features. We also propose new distance measures based on KL distance and $\mathcal{L}_2$-norm, which are proven to be as successful as the commonly used KL distance with a much less computational burden. Experiments establish the superiority of our novel covariance descriptor based on complex wavelets and new distance measures over the existing eigenvalue-based method [1]. For future work, we will try to find a representative training covariance matrix on a Riemannian manifold for each class in order to avoid computing distances to all matrices. We are also planning to devise a single-tree transform that mimics the behavior of DT-CWT.

REFERENCES

[1] A. Suhre, T. Ersahin, R. Cetin-Atalay, and A. E. Cetin, "Microscopic image classification using sparsity in a transform domain and bayesian learning," in *19th European Signal Processing Conference*, 2011, pp. 1005–1009.

[2] I. W. Selesnick and K. Y. Li, "Video denoising using 2d and 3d dual-tree complex wavelet transforms," in *Wavelet Appl Signal Image Proc. X (Proc. SPIE 5207*, 2003, pp. 607–618.

[3] P. Loo and N. Kingsbury, "Digital watermarking using complex wavelets," in *Image Processing, 2000. International Conference on*, vol. 3, 2000, pp. 29 –32.

[4] G. Chen, T. Bui, and A. Krzyzak, "Palmprint classification using dual-tree complex wavelets," in *Image Processing, 2006 IEEE International Conference on*, Oct 2006, pp. 2645 –2648.

[5] M. Thamarai and R. Shanmugalakshmi, "Video coding technique using swarm intelligence in 3-d dual tree complex wavelet transform," in *Machine Learning and Computing (ICMLC), 2010 Second International Conference on*, Feb 2010, pp. 174 –178.

[6] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 123 – 151, Nov. 2005.

[7] I. Selesnick, "Hilbert transform pairs of wavelet bases," *Signal Processing Letters, IEEE*, vol. 8, no. 6, pp. 170 –173, June 2001.

[8] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Computer Vision ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin / Heidelberg, 2006, vol. 3952, pp. 589–600.

[9] N. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, vol. 357, pp. 2543–2560, 1997.

[10] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 1713–1727, 2008.

[11] Y. Habiboglu, O. Gunay, and A. Cetin, "Flame detection method in video using covariance descriptors," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 1817 –1820.

[12] H. Tuna, I. Onaran, and A. Cetin, "Image description using a multiplier-less operator," *Signal Processing Letters, IEEE*, vol. 16, no. 9, pp. 751 –753, sept. 2009.

[13] R. Maree, P. Geurts, J. Piater, and L. Wehenkel, "Random subwindows for robust image classification," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, June 2005, pp. 34 – 40 vol. 1.

[14] Y. Shinohara, T. Masuko, and M. Akamine, "Covariance clustering on riemannian manifolds for acoustic model compression," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 4326 –4329.

[15] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226 –239, Mar 1998.