Project Title:	MIRACLE - Microscopic Image Processing Analysis Coding and Modelling Environment
Contract No:	PIRSES-GA-2009-247091
Instrument:	SESAM
Thematic Priority:	Medical image processing
Start of project:	1 May 2010
Duration:	36 months

Deliverable No: D5

WP1 Final Report

Due date of deliverable:	1 March 2011
Actual submission date:	26 March 2013
Version:	1
Main Authors:	Alexander Suhre, Enis Cetin,

Project ref. number	PIRSES-247091
Project title	MIRACLE - Microscopic Image Processing Analysis Coding and Modelling Environment

Deliverable title	Co-difference method for microscopic images
Deliverable number	D2
Deliverable version	V1
Previous version(s)	-
Contractual date of delivery	1 June 2011
Actual date of delivery	26 March 2013
Deliverable filename	MiracleDeliverable5WP1.doc
Nature of deliverable	Report
Dissemination level	Public
Number of pages	5
Workpackage	1
Partner responsible	BILKENT
Author(s)	Alexander Suhre, Enis Cetin, Furkan Keskin
Editor	Alexander Suhre
EC Project Officer	Alexandra Pedersen

Abstract	Final Report on the scientific work carried out in WP1 during the exchange program.
Keywords	Feature extraction, microscopic images

Current Methods of Feature Extraction for Microscopic Images

The widespread use of Computer-assisted diagnosis (CAD) can be traced back to the emergence of digital mammography in the early 1990's. Recently, CAD has become a part of routine clinical detection of breast cancer on mammograms at many screening sites and hospitals in the United States. In fact, CAD has become one of the major research subjects in medical imaging and diagnostic radiology. Given recent advances in high-throughput tissue bank and archiving of digitized histological studies, it is now possible to use histological tissue patterns with computer-aided image analysis to facilitate disease classification. There is also a pressing need for CAD to relieve the workload on pathologists by sieving out obviously benign areas, so that pathologist can focus on the more difficult-to-diagnose suspicious cases. For example, approximately 80% of the 1 million prostate biopsies performed in the US every year are benign; this suggests that prostate pathologists are spending 80% of their time sieving through benign tissue.

Researchers both in the image analysis and pathology fields have recognized the importance of quantitative analysis of pathology images. Since most current pathology diagnosis is based on the subjective (but educated) opinion of pathologists, there is clearly a need for quantitative image-based assessment of digital pathology slides. This quantitative analysis of digital pathology is important not only from a diagnostic perspective, but also in order to understand the underlying reasons for a specific diagnosis being rendered (e.g., specific chromatin texture in the cancerous nuclei which may indicate certain genetic abnormalities). In addition, quantitative characterization of pathology imagery is important not only for clinical applications (e.g., to reduce/eliminate inter- and intra-observer variations in diagnosis) but also for research applications (e.g., to understand the biological mechanisms of the disease process).

A large focus of pathological image analysis has been on the automated analysis of cytology imagery. Since cytology imagery often results from the least invasive biopsies (e.g., the cervical Pap smear), they are some of the most commonly encountered imagery for both disease screening and biopsy purposes. Additionally, the characteristics of cytology imagery, namely the presence of isolated cells and cell clusters in the images and the absence of more complicated structures such as glands make it easier to analyze these specimens compared to histopathology. For example, the segmentation of individual cells or nuclei is a relatively easier process in such imagery since most of the cells are inherently separated from each other. Histopathology slides, on the other hand, provide a more comprehensive view of disease and its effect on tissues, since the preparation process preserves the underlying tissue architecture. As such, some disease characteristics, e.g., lymphocytic infiltration of cancer, may be deduced only from a histopathology image. Additionally, the diagnosis from a histopathology image remains the 'gold standard' in diagnosing considerable number of diseases including almost all types of cancer . The additional structure in these images, while providing a wealth of information, also presents a new set of challenges from an automated image analysis perspective. It is expected that the proper leverage of this spatial information will allow for more specific characterizations of the imagery from a diagnostic perspective. The analysis of histopathology imagery has generally followed directly from techniques used to analyze cytology imagery. In particular, certain characteristics of nuclei are hallmarks of cancerous conditions. Thus,

quantitative metrics for cancerous nuclei were developed to appropriately encompass the general observations of the experienced pathologist, and were tested on cytology imagery. These same metrics can also be applied to histopathological imagery, provided histological structures such as cell nuclei, glands, and lymphocytes have been adequately segmented (a complication due to the complex structure of histopathological imagery). The analysis of the spatial structure of histopathology imagery can be traced back to the works of Wiend *et al*., Bartels and Hamilton but has largely been overlooked perhaps due to the lack of computational resources and the relatively high cost of digital imaging equipment for pathology. However, spatial analysis of histopathology imagery has recently become the backbone of most automated histopathology image analysis techniques. Despite the progress made in this area thus far, this is still a large area of open research due to the variety of imaging methods and disease-specific characteristics.

Best wavelets for FL images

The dual-tree complex wavelet transform (DT-CWT) has been recently used in various signal and image processing applications. It has desirable properties such as shift invariance, directional selectivity and lack of aliasing. In the dual-tree CWT, two maximally decimated discrete 7 wavelet transforms are executed in parallel, where the wavelet functions of two different trees form an approximate Hilbert transform pair. Filterbanks for DT-CWT are shown in Figure 1. 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. 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.

With the 2-D dual-tree CWT, many ideas and techniques from Gabor analysis can be leveraged into wavelet-based image processing".

Microscopic cancer cell line images contain significant amount of oriented singularities. 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 for feature extraction from microscopic images. In the region covariance framework each pixel is mapped to a set of pixel properties which's covariances are measured and used as a 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.

2-D DT-CWT of an image is obtained by four real separable transforms. 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).

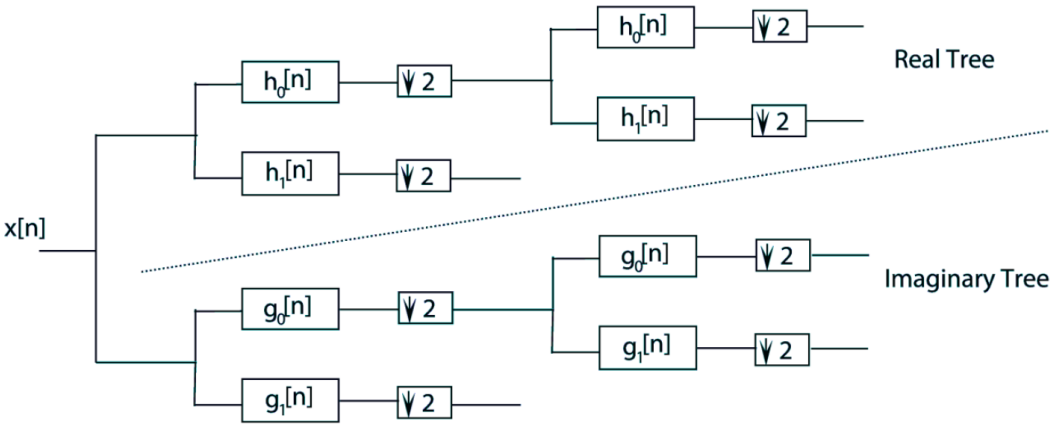

Figure 1: DT-CWT Filterbank

Covariance method for microscopic images

Cancer cell lines are widely used for research purposes in laboratories all over the world. Computerassisted classification of cancer cells can alleviate the burden of manual labeling and help cancer research. In this workpackage, we presented a novel computerized method for cancer cell line image classification.

The aim is to automatically classify 14 different classes of cell lines including 7 classes of breast and 7 classes of liver cancer cells. Microscopic images containing irregular carcinoma cell patterns are represented by subwindows which correspond to foreground pixels. For each subwindow, a covariance descriptor utilizing the dual-tree complex wavelet transform (DT-CWT) coefficients and several morphological attributes are computed. Directionally selective DT-CWT feature parameters are preferred primarily because of their ability to characterize edges at multiple orientations which is the characteristic feature of carcinoma cell line images. A Support Vector Machine (SVM) classifier with radial basis function (RBF) kernel is employed for final classification. Over a dataset of 840 images, we achieve an accuracy above 98%, which outperforms the classical covariance-based methods. The proposed system can be used as a reliable decision maker for laboratory studies. Our tool provides an automated, time- and cost-efficient analysis of cancer cell morphology to classify different cancer cell lines using image-processing techniques, which can be used as an alternative to the costly short tandem repeat (STR) analysis. The data set used in this manuscript is available as supplementary material through *http://signal.ee.bilkent.edu.tr/cancerCellLineClassi_cationSampleImages.html*.

Co-difference method for microscopic images

A new framework for signal processing is introduced based on a novel vector product definition that permits a multiplier- free implementation. First a new product of two real numbers is defined as the sum of their absolute values, with the sign determined by product of the hard-limited numbers. This new product of real numbers is used to define a similar product of vectors in $\mathbb{R}^N$. The new vector product of two identical vectors reduces to a scaled version of the l_1 norm of the vector. The main advantage of this framework is that it yields multiplication-free computationally efficient algorithms for performing some important tasks in signal processing. An application to the problem of cancer cell line image classification is presented that uses the notion of a co-difference matrix that is analogous to a covariance matrix except that the vector products are based on our new proposed framework. Results show the effectiveness of this approach when the proposed co-difference matrix is compared with a covariance matrix.