FRAMES IN BIOIMAGING

Amina Chebira[1] *and Jelena Kovačević*[1,2]

[1] Dept. of BME and Center for Bioimage Informatics, [2] Dept. of ECE
Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

We survey our work on adaptive multiresolution (MR) approaches to the classification of biological images. The system adds MR decomposition in front of a generic classifier consisting of feature computation and classification in each MR subspace, yielding local decisions, which are then combined into a global decision using a weighting algorithm. The system tested on different datasets (subcellular protein location images, drosophila embryo images and histological images images) gave very high accuracies. We hypothesize that the space-frequency localized information in the multiresolution subspaces adds significantly to the discriminative power of the system. Moreover, we show that a vastly reduced set of features is sufficient. Finally, we prove that frames are the class of MR techniques that performs the best in this context. This leads us to consider the construction of a new family of frames for classification, which we term lapped tight frame transforms.

1. BACKGROUND AND MOTIVATION

Systems biology entails the study of the interactions between the components of a biological system and the mechanisms by which these interactions give rise to the function and behavior of that system. Such approach encompasses mathematical and computational modeling based on quantitative data collected within each component of the biological system. Advances in biochemistry, probes, and microscopy gave the biologists the opportunity to observe cells and cell processes at a level never seen before, which led to the collection of huge amounts of 2D, 3D and even higher-dimensional data. As a result, visual inspection of these datasets, always error-prone, nonreproducible and subjective, became impractical as well. Hence the need for automated, accurate and efficient systems to extract knowledge contained in the collected data.

Automated knowledge extraction requires the expertise developed in signal processing, machine learning and mathematics. In the project of determination of protein subcellular location patterns, Murphy et al. identified classification as the underlying problem [1]. Similarly, in the project of determination of developmental stages in fly embryos [2], we realized that the problem is again that of classification. Not surprisingly then, in several other projects, such as the development of teratomas in stem cells, the need for classification emerged. Thus, an accurate and efficient algorithm for classification would be of great use to biologists, motivating the developments in this work.

Classification. Assume that the images are of size $N \times N$ and let $\mathbb{R}$ denote the set of intensities covered by all the images in the given dataset. The classification problem can be formulated as designing a map from the *signal space* of the examined images $\mathcal{X} \subset \mathbb{R}^{N \times N}$, to a *response space* $\mathcal{Y} \subseteq \{1, 2, ..., C\}$ of class labels. The decision d is the map, $d\colon \mathcal{X} \mapsto \mathcal{Y}$ that associates an input image with a class label. To reduce the dimensionality of the problem, one sets up a feature space $\mathcal{F} \subset \mathbb{R}^f, f \leq N^2$, between the input space and the response space. The feature extractor θ is the map $\theta\colon \mathcal{X} \mapsto \mathcal{F}$, and the classifier ν is the map $\nu\colon \mathcal{F} \mapsto \mathcal{Y}$. The goal is to find a (θ, ν) pair that maximizes the classification accuracy.

Features are numerical descriptors that characterize the input data, usually in a lower-dimensional space. We focus on the following feature sets: (a) *Haralick Texture Features (T_1 with 13 features or T_2 with 26 features)*, are calculated using four co-occurrence matrices which are combined in various ways to give either 13 or 26 measures. (b) *Morphological Features (M, 16 features)*, visually describe distinctive aspects of images as discerned by the human eye. (c) *Zernike Moment Features (Z, 49 features)*, computed for an image are similarity measures between the corresponding Zernike polynomials and the image. We focus on a specific class of nonlinear classifiers: Neural networks (NNs), based on grouping the input vectors (features) into intersections of clusters of one type while the union of all such intersections yields the entire feature space.

In the problem of determination of protein subcellular location patterns, the heart is a set of features (T_1, T_2, M, and Z) describing the spatial distribution of proteins in each cell image. Of particular relevance to the work described here is the use of simplest MR features such as wavelet (30 features) and Gabor (60 features) features, as the addition of these resulted in a significant improvement of classification accuracy to 91.5% for the 2D HeLa dataset [1]. As the introduction of the simplest MR features produced a statistically significant jump in classification accuracy, our hypothesis was that more sophisticated MR techniques would result in even more accurate classification.

Multiresolution Techniques. MR techniques have been extensively studied and used in signal and image processing over the past two decades [3]. MR processing means analysis and processing of data at different resolutions and/or scales. MR transforms decompose a signal into zooming spaces (coarse spaces and many detail spaces called *subbands*) and are implemented by *filter banks (FBs)*, through filtering and sampling.

Nonredundant Multiresolution Techniques: Bases. Most of MR techniques in use are nonredundant— the underlying mathematical structures are *bases (MRBs)*. Assuming $\mathbb{R}^m$ or $\mathbb{C}^m$, a basis for such a space, $\Psi = \{\psi_i\}_{i=0}^{m-1}$, we associate to it a matrix (operator) which we will also call Ψ, and which has basis vectors as its columns ($\psi_{i,j}$ is the jth element of the ith basis vector). Given a pair of biorthogonal bases $(\Psi, \tilde{\Psi})$ dual to each other, a signal x belonging to $\mathbb{R}^m$ or $\mathbb{C}^m$ can be expressed as:

$$x = \Psi X = \Psi \tilde{\Psi}^* x, \quad (1)$$

This work was supported in part by NSF grants CCF-0515152 and EF-0331657, the PA State Tobacco Settlement, Kamlet-Smith Bioinformatics Grant.

where X is the vector from $\mathbb{R}^m$ or $\mathbb{C}^m$ of so-called *transform* coefficients (inner products), and $\tilde{\Psi}^*$ denotes the Hermitian transpose of the dual basis $\tilde{\Psi}$. If the expansion is into an orthonormal basis (ONB), then $\Psi = \tilde{\Psi}$ and the above becomes $\Psi\Psi^* = I$, which further implies that Ψ is a unitary matrix.

The only infinite-dimensional class of MR decompositions we discuss here are those implemented by FBs, as these are bases used in applications. The vectors (signals) live in the infinite-dimensional Hilbert space $l_2(\mathbb{Z})$[1]. When $m = n$, we deal with critically-sampled FBs implementing bases. A FB decomposition can be expressed as in (1) where x is now an infinite sequence belonging to $l_2(\mathbb{Z})$, X is an infinite sequence of transform coefficients (inner products) in $l_1(\mathbb{Z})$, and Ψ is the basis expansion matrix given in a setting with finite impulse response (FIR) filters. The matrix Ψ is used in the synthesis FB (the reconstruction step) whereas its dual, $\tilde{\Psi}$, is used in the analysis FB (the decomposition step). Assume that the nonzero support of the filter ψ_i, or, its length is $l = km$ (if not, we can always pad with zeros), and write the basis operator as a block circulant matrix with blocks Ψ_r is of size $m \times m$, containing filter coefficients. We can rephrase the basis decomposition in the z-domain as well using polyphase analysis. A polyphase matrix $\Psi_p(z)$ collects the subsequences modulo n. For bases, $\Psi_p(z)$ is of size $m \times m$:

$$\Psi_p(z) \;=\; \sum_{r=0}^{k-1} \Psi_r z^{-r}. \tag{2}$$

A paraunitary polyphase matrix (representing an ONB) satisfies $\Psi_p(z)$ $\Psi_p^*(z) = \mathrm{I}$, where c is a constant.

When the filter length l is equal to the sampling factor m, we have a *block transform*. Then, only Ψ_0 is nonzero, making Ψ block-diagonal. In effect, since there is no overlap between processed blocks, this can be analyzed as a finite-dimensional case, where both the input and the output are m-dimensional vectors. This shows how finite-dimensional bases can be analyzed in the FB context.

In practice, the use of block transforms can produce artifacts known as "blocking effects" (since there is no overlap between the basis functions—processed blocks), and thus solutions were sought with longer basis functions. One such solution is the *Lapped Orthogonal Transform (LOT)*. The LOTs can be seen as a class of m-channel FBs implementing bases, originally developed for filters of length $l = 2m$ and later generalized to arbitrary integer multiples of m [4]. Compared to block transforms, the LOT keeps the same number of filters but doubles their length, which means that the basis functions of adjacent blocks overlap by half their size, thus removing the blocking effects. However, LOTs are not solely determined by their length, but by the specific form of their basis vectors as well.

In general, for a FB with filter length $l = 2m$, the time-domain matrix Ψ has a double diagonal, that is, only Ψ_0 and Ψ_1 exist. Thus, (2) reduces to

$$\Psi_p(z) \;=\; \Psi_0 + z^{-1}\Psi_1, \tag{3}$$

where $\Psi_r, r = 0, 1$, are $m \times m$ matrices with $(\Psi_r)_{j,i} = \psi_{i,j}$ for $i = 0, \ldots, m-1$ and $j = mr, \ldots mr + m - 1$. Since the LOT is a unitary transform, that is, $\Psi\Psi^* = \Psi^*\Psi = I$ the following must be satisfied:

$$\Psi_0\Psi_0^* + \Psi_1\Psi_1^* \;=\; \Psi_0^*\Psi_0 + \Psi_1^*\Psi_1 = I, \tag{4}$$

$$\Psi_0^*\Psi_1 = \Psi_1^*\Psi_0 = \;0, \quad \Psi_0\Psi_1^* = \Psi_1\Psi_0^* = 0. \tag{5}$$

Two main classes of LOTs exist distinguished by whether they use cosines or complex exponentials in their basis functions. We concentrate here on a particular family that uses cosine basis functions: The The Princen-Johnson-Bradley LOT (PJB-LOT) [5] basis functions are given by:

$$\psi_{i,j} \;=\; \sqrt{\frac{1}{m}}\cos\left(\frac{\pi(2i+1)(2j-m+1)}{4m}\right), \tag{6}$$

with $i = 0, \ldots, m-1$, and $j = 0, \ldots, 2m-1$.

With this construction, we will have, similarly to the Discrete Fourier Transform (DFT), fixed basis functions allowing no freedom in design, which can be remedied by adding a window that multiplies each filter resulting in a modulated FB over the frequency band. This modulated FB can be modeled as $W\Psi$, where the window $W = \mathrm{diag}\{w_j\}_{j=0}^{2m-1}$ is symmetric $w_j = w_{2m-1-j}, j = 0, \ldots 2m-1$ and has to satisfy $w_j^2 + w_{m-1-j}^2 = 2$, for $j = 0, \ldots, m-1$.

The *Discrete Wavelet Transform (DWT)*, a famous MR tool, is a basis expansion and as such nonredundant (critically sampled). The *dyadic* DWT is built by iterating a two-channel FB with sampling factor $n = m = 2$ on the lowpass channel. One can also build trees by, at each level, iterating on any subset of the branches of the FB. This is knows as *Wavelet Packets (WP)*. Both DWTs and WPs can be block transform or not depending on the length of the filters.

In our previous work on fingerprint images [6], we used an adaptive wavelet packet (WP) approach in combination with correlation filters to solve the recognition problem, with considerable success. In the same work, we found the system was sensitive to shift variance introduced by wavelet bases, prompting us to turn to redundant ones—frames.

Redundant Multiresolution Techniques: Frames. In $\mathbb{R}^n$ or $\mathbb{C}^n$, a frame is defined as a set Φ of m frame vectors $\Phi = \{\varphi_0, \ldots, \varphi_{m-1}\}$ where m is larger than n. As for bases, we associate to the frame a rectangular matrix of size $n \times m$, also called Φ, that has the frame vectors as its columns.

Similarly to bases, one can check that frames expand signals in $\mathbb{R}^n$ with $x = \Phi X = \Phi\tilde{\Phi}^* x$, where $\tilde{\Phi}$ represents the dual frame. If $\Phi = \tilde{\Phi}$, then we have what is called a *tight frame (TF)*, and the expansion becomes $\Phi\Phi^* = I$. If all frame vectors have the same norm, the frame is termed an *equal-norm frame (ENF)*, while if all the norms are equal to 1, this is a *unit-norm frame (UNF)*. By combining this with the requirement of tightness, we can have *equal-norm tight frames (ENTFs)*, as well as *unit-norm tight frames (UNTFs)*. An introductory account on frames is given in [7].

In an m-channel FB with sampling factor n, if $m > n$, we deal with an oversampled FB implementing a frame. For a TF, $\tilde{\varphi}_i = \varphi_i$. The FB frame decomposition can be expressed as in (1) (substituting Φ for Ψ), where x is an infinite sequence belonging to $l_2(\mathbb{Z})$, X is an infinite sequence of transform coefficients (inner products), and Φ is the frame expansion matrix. Assuming again that the nonzero support of the filters (frame vectors) length is $l = kn$, we can write the frame operator Φ as for bases, with matrices $\Phi_r, r = 0, \ldots, k-1$, being rectangular of size $n \times m$.

We can rephrase the frame decomposition in the z-domain as well, where a FB implements a TF decomposition in $l_2(\mathbb{Z})$ *if and only if* its polyphase matrix $\Phi_p(z)$ is paraunitary [8]. For frames, the polyphase matrix $\Phi_p(z)$ is of size $n \times m$ and can be written as in (2) (substituting Φ for Ψ), where Φ_r are of size $n \times m$

In an ever-continuing search for new frame families, an appealing option is the process of obtaining TFs from ONBs in larger dimensions, known as the Naimark Theorem or *seeding* [9].

Definition 1 *We say that a frame Φ is obtained by seeding from a basis Ψ by deleting a suitable set of columns of Ψ. We write $\Phi^* =$*

[1] In fact, we can investigate finite-dimensional MR decompositions within the FB framework as well.

$\Psi[\mathbb{J}]$ *where* $\mathbb{J}$ *is the index set of the retained columns.*

All tight frames can be obtained this way. One of the most famous frame families, the Harmonic Tight Frames (HTFs) is obtained by seeding the DFT. In FB parlance, seeding is done on the polyphase matrix. Given $\Psi_p(z)$, the $m \times m$ polyphase matrix associated with a basis of size m, then $\Psi_p(z) = \Psi_0$, and

$$\Phi_p^*(z) = \Phi_0^* = \Psi_p[\mathbb{J}] \tag{7}$$

is the transpose of the frame polyphase matrix. When $l = n$, that is, the length of the frame vectors equals the sampling factor, we obtain a block transform, such as HTFs.

2. MULTIRESOLUTION CLASSIFICATION ALGORITHM

In the last section, we saw that the classification problem is ubiquitous in biology, and that MR techniques might make classification more accurate. The results obtained in [6] seem to indicate that adaptive MR techniques, frames in particular, might be needed. Having motivated the use of adaptive MR in classification as well as the need for redundant MR transforms, we now test that hypothesis. We first describe the adaptive MR classification algorithm developed; details and results are described in [10, 11, 12, 13].

Main Idea. As argued earlier, we would like to extract discriminative features within space-frequency localized subspaces. These are obtained by MR decomposition; that is, instead of adding MR features as in [1], we compute features in the MR-decomposed subspaces. Our initial idea was to use WPs since they adapt themselves

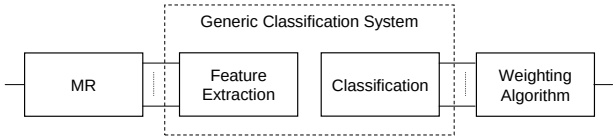


Fig. 1. Adaptive MR classification system [10].

to the signal at hand, and just as in the fingerprint case, prove that adaptivity significantly improves the recognition system. So, ideally, we would characterize each class by the best wavelet packet tree that represents it. However, this is possible only if a suitable cost function can be found. Given that we have no natural cost function available, we decided to mimic a wavelet-packet like system by adding a weighting procedure at the end of our system, allowing us to weigh the decisions of each subband in a fully grown tree. This way, a very low weight emulates a pruned branch in the tree. Thus, we propose a system with an MR decomposition block in front (see Fig. 1), followed by feature computation and classification in each of the subspaces, which are then combined through a weighting process, providing adaptivity.

Multiresolution Block. In our classification system, any MR transform can be used. In particular, amongst the MRBs, we used the DWT, DFT, Discrete Cosine Transform (DCT) and others, while amongst the MRFs, we used the Double-Density DWT (DD-DWT), Dual-Tree Complex Wavelet Transform (DT-CWT) and the Stationary Wavelet Transform (SWT), which is the most redundant transform. Note that here, we use all the subbands of the decomposition tree, not only the leaves. Thus, it might be abuse of language to call a transform a DWT. For example, for 2 levels, we have a total of $S = 21$ subbands (original image + 4 subbands at the first level + 16 subbands at the second level).

Feature Extraction and Classifier. We start with the feature sets used in [1]: Haralick texture features (set T_1, 13 features), morphological (set M, 16 features) and Zernike moments (set Z, 49 features). Unlike in [1], we do not use wavelet/Gabor features because the MR advantage given by these will be achieved by our MR decomposition. Therefore, our total number of features is 78, as opposed to 174 in [1].

Instead of combining all features into a single feature vector, we allow each feature set its own feature vector per subband effectively bringing the number of subbands to $3 \cdot S = 63$ when using two levels of decomposition and all three feature sets. Note that although we have decreased the number of features significantly, we have also increased the number of classifiers, because we now have one classifier per subband. Evaluating this computational trade-off is a task for future work.

New Texture Feature Set T_3. As we will show later on, the Haralick texture features seem to possess the most discriminative power, so we looked more closely into these. We changed the way that Haralick combines the initial four sets of features. We note that horizontal and vertical co-occurrence matrices are fundamentally different from the diagonal ones because adjacent neighboring pixels are spatially closer than diagonal neighboring pixels. Therefore, instead of averaging the features from all four sets, we create our first set of 13 features by averaging horizontal and vertical measures, and a second set of 13 features by averaging diagonal measures. Thus, we end up with a new feature set T_3 of 26 features [10].

Neural Networks. We decided to use a two-layer NN classifier. The first layer contains a node for each of the input features, each node using the Tan-Sigmoid transfer function. The second layer contains a node for each output and uses a linear transfer function (no hidden layers are used). In our design, when training, each output from the second layer corresponds to a class, and each training image will have an output of 1 for the class of which it is a member and a 0 for all other classes.

Weighting Procedure. Fig. 1 shows a graphical representation of a generic MR classification system, including the process of combining all of the subband decisions into one. We use weights for each subband to adjust the importance that a particular subband has on the overall decision made by the classification system. If the weights are chosen such that the no-decomposition weight is equal to 1, and all other weights are 0, we will achieve the same output vector as we would have without using the adaptive MR system. Therefore, we know that there exists a weight combination that will do at least as well as the generic classifier (when no MR is involved) in the training phase. Our goal is to decide how to find the weight vector that achieves the highest overall classification accuracy on a given dataset. We developed three versions of the weighting algorithm: open-form (OF), per-dataset closed-form (PD-CF))and per class closed-form (PC-CF). The PD-CF algorithm assigns one weight vector for the entire dataset, whereas the PC-CF assigns a weight vector for each class in the dataset. The latter goes back to our original idea of having a wavelet packet tree characterizing each class, only in this case, we do not necessarily obtain a tree.

The NN block outputs a series of decision vectors for each subband of each training image. Each decision vector $d_s^{(r)}$ contains C numbers (where C denotes number of classes) that correspond to the "local" decisions made by the subband s for a specific image r.

The classifier is evaluated using nested cross validations (five-fold cross validation in the NN block and ten-fold during the weighting process). One problem with this technique is that the initial ordering of the images determines which images are grouped together for training and testing in each fold of the cross validation. We solve

this problem by running multiple trials, each with a random initial ordering of the images.

Open-Form Algorithm (OF). If using the OF algorithm, we initialize all the weights, and a global decision vector is computed through an iterative procedure using a weighted sum of the local decisions [10].

Per-Dataset Closed-Form Algorithm (PD-CF). The CF solution does not use an iterative algorithm; rather, it finds the weight vector by solving a minimization problem in the least-square sense.

Assume we have R training images. For each training image $r = 1, \ldots, R$, the vector $d_s^{(r)} = (d_{s,c}^{(r)})^T$ for classes $c = 1, \ldots, C$, is the $C \times 1$ decision vector at the output of each subband classifier s, where $d_{s,c}^{(r)}$ indicates the confidence of subband s that the training image r belongs to class c. For each training image r, the weighting block takes as input the subband (local) decision vectors $d_s^{(r)}$ and combines them into a single output decision vector as follows:

$$\sum_{s=1}^{S} w_s d_s^{(r)}. \tag{8}$$

We can rewrite the above by, for each training image r, forming a matrix $D^{(r)}$ of size $C \times S$, where each element $D_{c,s}^{(r)}$ is the value at position c of the decision vector $d_s^{(r)}$ of subband classifier s. We can now compute $D^{(r)}w$, where $w = (w_1, \ldots, w_S)^T$ is of size $S \times 1$. Thus, we want to find a weight vector w common to all training images $r = 1, \ldots, R$.

A possible solution for w is the one that minimizes the error in the least-square sense:

$$w_{win} \;=\; \arg\min_{w} \sum_{r=1}^{R} \|d^{(r)} - D^{(r)}w\|^2, \tag{9}$$

where $d^{(r)}$ is the desired target decision vector of size $C \times 1$, with 1 in the position of the true class, and 0 elsewhere. We need to rewrite the above in a direct error-minimization form. We thus define a target output vector d of size $CR \times 1$, as a vector which concatenates all the target decision vectors $d^{(r)}$ as follows:

$$d \;=\; \left(\underbrace{d_1^{(1)}, d_2^{(1)}, \ldots, d_C^{(1)}}_{\text{training image 1}}, \ldots, \underbrace{d_1^{(R)}, \ldots, d_C^{(R)}}_{\text{training image } R} \right)^T,$$

and a $CR \times S$ matrix D consisting of the all the decision matrices $D^{(r)}$ of all the training data stacked on top of each other. That is, $(D)_{cr,s} \;=\; D_{c,s}^{(r)}$ for $c = 1, \ldots, C, r = 1, \ldots, R$ and $s = 1, \ldots, S$. We can now rewrite (9) as:

$$w_{win} \;=\; \arg\min_{w} \|d - Dw\|, \tag{10}$$

which possesses a CF solution and can be computed efficiently.

Then, for a testing image t, we compute its decision vector $\delta = (\delta_1, \delta_2, \ldots, \delta_C)$ as follows:

$$\delta \;=\; \sum_{s=1}^{S} w_{win,s} d_s^{(t)},$$

where $d_s^{(t)}$ are the local decision vectors for t. The classification decision is then made as $c_{win} \;=\; \arg\max_c \delta_c$, that is, the winning class corresponds to the index of the highest coefficient in δ.

Per-Class Closed-Form Algorithm (PC-CF). To make the system truly adaptive, it is reasonable to assume that different classes require different weight vectors. Thus, we propose a system where, instead of a single weight vector w for the whole training dataset, each class c has its own weight vector w_c [13]. As opposed to (8), the entries in the output decision vector are now computed as:

$$D^{(r)}w_c, \qquad c = 1, \ldots, C. \tag{11}$$

Now, the weights can be grouped together to form an $S \times C$ matrix W so that each column represents a class-specific weight vector. Equation (11) can be rewritten as $\mathrm{diag}\left(D^{(r)}W\right)$. Recall that $D^{(r)}$ is of size $C \times S$ and thus d is of size $C \times C$ (compare this to (8)). To learn these weights, we again use the training set and look for a solution that minimizes the squared error:

$$W_{win} \;=\; \arg\min_{W} \sum_{r=1}^{R} \|d^{(r)} - \mathrm{diag}\left(D^{(r)}W^{(r)}\right)\|^2. \tag{12}$$

To obtain an expression analogous to (10) and be able to apply standard methods, we have to define v as the vector concatenating all class-specific weight vectors:

$$v \;=\; \left(W_{1,1}, W_{1,2} \ldots W_{1,C}, \ldots, W_{S,1}, \ldots, W_{S,C}\right)^{\top}. \tag{13}$$

We now define D as the $CR \times CS$ block matrix $(D)_{rc,sc} \;=\; d_c^{(r)}$, where $d_c^{(r)}$ is the vector $(D_{c,1}^{(r)}, D_{c,2}^{(r)}, \ldots, D_{c,S}^{(r)})$ and $r = 1, \ldots, R$. We can now write a minimization problem equivalent to the one in (12), and which we can solve using standard techniques:

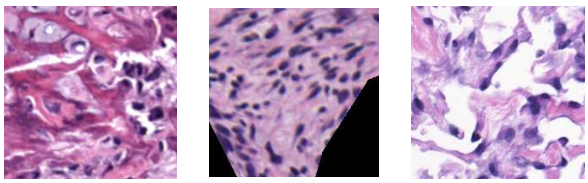
$$v_{win} \;=\; \arg\min_{v} \|d - Dv\|. \tag{14}$$

Decomposition Tree Pruning. Our long-term goal in developing an adaptive MR classification system was to find a WP-like decomposition, where each class would induce a different MR subtree. While the authors have done just that in [6], we needed a cost function which is specific to the dataset used. Our goal is thus to have a more generic system and to achieve a WP-like system but without the need for a cost function. We come close to this goal here, where we identify the set of discriminative subbands for each class (not necessarily a subtree). Once the weight vectors are computed (using any version of the CF weighting algorithm), we use the values of the weights to regulate the MR decomposition. In particular, subbands with low weight can be pruned as long as the remaining subbands are still sufficient to classify the image correctly. This way, the pruned subbands and their associated features need not be computed, resulting in computational savings (although not increased accuracy). We propose to keep the high-weight subbands, so that at least a certain ratio, defined as the fraction of the sum of kept weights over the sum of all the weights, of subbands are kept.

3. CLASSIFICATION PROBLEMS IN BIOIMAGING

While we have developed the current algorithm by learning from each application as we went along, we decided to first present all algorithmic accomplishments and then discuss results in various application domains. We do that now and use different instantiations of the MR classification algorithm depending on the dataset at hand.

Determination of Protein Subcellular Location Patterns. To evaluate our MR approach, we use the 2D HeLa set depicting PSL [1]. The proteins in the dataset were labeled using immunofluorescence,



(a) (b) (c)

Fig. 2. Sample H&E-stained images from three of the six tissue class: (a) bone, (b) mesenchyme (embryonic connective tissue) and (c) myenteric plexus.

and thus, we know the ground truth, that is, which protein was labeled in each cell and subsequently imaged. The details of our results in this area can be found in [10]. The challenge in this dataset is that images from the same class may look different while those from different classes may look very similar. The dataset is publicly available (http://murphylab.web.cmu.edu) and contains 50 single-cell images of size 512×512, in each of $C = 10$ classes. The 10 classes of subcellular location patterns were obtained by labeling an endoplasmic reticulum protein, two Golgi proteins (giantin and gpp130), a lysosomal protein, a mitochondrial protein, a nucleolar protein, two cytoskeletal proteins (actin and tubulin), an endosomal protein, and DNA. The best previously described overall classification accuracy on this dataset is 91.5% [1]. We achieve the best classification accuracy of 95.40% using MRFs [10].

Detection of Developmental Stages in Drosophila Embryos. The details of our results in this area can be found in [11]. The dataset consists of 60 time-lapse, fluorescence microscopy z-stacks (3D volumes in time) of developmental stages of *Drosophila* embryos. The stacks are acquired roughly every 10 minutes. The number of slices per stack varies; it is 5 slices for normal sets and 7 slices for delayed/abnormal. The number of time points is typically 15 for normal/abnormal and around 30 for delayed. All the slices have been tagged by a human expert so we have reliable ground truth. The highest classification accuracy achieved by our system is 93.35% [11].

Classification of Histological Stem-Cell Teratomas. We start with H&E images (at 10x magnification, see Fig.2) that depict sample images from multiple tissues (classes) contained in the teratomas. We use six classes for this experiment: mesenchyme (embryonic connective tissue), skin, myenteric plexus, bone, necrotic (dying or dead tissue), and striated muscle. The best classification accuracy of 87.72% is achieved by MRFs [12, 14].

Results and Discussion. By observing the results for all applications, we drew the following conclusions:

- In all cases, MR significantly outperforms no MR, thus demonstrating that classifying in MR subspaces indeed improves classification accuracy.
- The redundant transforms (MR frames) do better than nonredundant ones (MR bases). In all cases, the SWT achieves the best classification accuracy.
- While a slightly higher classification accuracy is obtained by using all three feature sets as well as both T and M, the larger number of features and additional complexity of using M and Z features do not justify the slight improvement in accuracy This "flat" trend is good news as we can use a significantly reduced feature set and still obtain a fairly high classification accuracy.
- The closed-form version of the weighting process gives slightly better results than the open-form.
- As expected, when used, while pruning does not improve the accuracy of the system, it does make the system more efficient.
- In general, the class-adaptive method seems to do better than the dataset-adaptive one.

4. DESIGN OF NEW FRAME FAMILIES

Examining the results we obtained with our MR classification system, we found the trends to be similar: MR significantly outperforms NMR and the best results are invariably obtained by frames. Taking into consideration the computational cost, it is important to have a system that is efficient in addition to being accurate. The SWT is the most accurate here but also the most redundant. Therefore, to allow for a trade-off between accuracy and cost, we would like to create new redundant transforms that are less redundant but still afford very good accuracies when it comes to classification.

The question now is: How do we go about constructing new families and what do we look for? Most of the known frame families (though not all) are block ones (finite dimensional) leading to blocking effects. We want to have efficient implementations as well as be able to flexibly decide on the requisite amount of redundancy. These requirements made us think of LOTs, which, on top of being computationally efficient, have the advantage of processing blocks of overlapping data and hence eliminate blocking artifacts. So the question is: Could we construct a similar transform with frames? Our idea is to seed LOTs to obtain a new class of frames we name *Lapped Tight Frame Transforms (LTFTs)*. That is, we want to find filter-bank frames seeded from the LOTs in the hope they will inherit all the good properties LOTs possess. Obtained by seeding, the LTFTs could thus be seen both as the frame counterpart of LOT bases as well as the infinite-dimensional, filter-bank counterpart of the most famous frame family—Harmonic Tight Frames (HTFs, seeded from the DFT). Details can be found in [15].

Lapped Tight Frame Transforms.. We propose a host of new frame families we denote as *Lapped Tight Frame Transforms (LTFTs)* (details about this part of our work are given in [15]):

Definition 2 *A lapped tight frame is a frame transform (LTFT) seeded from an LOT.*

We start with $\Psi_p(z)$ being the $m \times m$ polyphase matrix associated with the LOT of size m. Then (3) holds and we seed the LOT to get:

$$\Phi_p^*(z) = \Phi_0^* + z^{-1}\Phi_1^* = \Psi_p[\mathbb{J}]. \tag{15}$$

The matrices Φ_r^* are now rectangular of size $n \times m$. For $r = 0, 1$, we have $(\Phi_r)_{i,j} = \psi^*_{i,mr+j}$, with $i = 0, \ldots, n-1$ and $j = 0, \ldots, m-1$. By Naimark Theorem, we know that this family is a TF, which implies that $\Phi_p(z)\Phi_p^*(z) = cI$. Note that as opposed to the LOT case, the matrix products do not commute anymore.

Princen-Johnson-Bradley LTFTs. All of the above is general and can be applied to any type of LOT. Let us now see through an example what happens when the obtained LTFT has been seeded by the PJB-LOT in (6). We will call these *Princen-Johnson-Bradley LTFTs (PJB-LTFTs)*. From the Naimark theorem, we know that any general seeding will result in a TF. However, if we want more than tightness (equal-norm, for instance), in general, there is no guarantee that we will obtain such a property. One needs to carefully choose the set of retained columns to preserve those properties. Indeed, seeding the DFT leads to HTFs only if the set of eliminated columns is contiguous. Choosing non contiguous columns leads to tight but not equal-norm frames. In the PJB case, it turns out that we can have a general seeding and any subset of columns (not necessarily contiguous) can be retained to obtain an ENTF. We summarize this result in the following (the proof is technical and not included here):

Lemma 1 *The PJB-LTFTs obtained by any seeding of the PJB-LOTs are ENTFs. That is, $\Phi_p^*(z) = \Psi_p[\mathbb{J}]$ for any subset $\mathbb{J}$ of columns is an ENTF, with norm $\sqrt{n/m}$.*

For PJB-LOTs we can compute $\mathrm{diag}(\Phi_0^*\Phi_0 + \Phi_1^*\Phi_1)_i = \|\varphi_i\|^2$. From Lemma 1, we know that without loss of generality we can assume that $\mathbb{J} = 0, \ldots, n-1$. Note that

$$\|\varphi_i\|^2 = \sum_{j=0}^{n-1} \psi^{*2}_{j,i} + \psi^{*2}_{j,i+m}.$$

In fact, we can find the equal norm as $\|\varphi_i\|^2 = n/m$, for $i = 0, \cdots, m-1$. That is, the LTFT obtained is equal-norm.

Window Design. If we start with the PJB-LOT with a window, and seed $W\Psi$, the TF obtained would loose its equal-norm property since $\|\varphi_i\|^2 = (n/m)w_i^2$. To preserve equal norm, we have to modulate directly the LTFT after seeding the LOT. In the PJB-LOTs, the window chosen was symmetric, that is, $w_i = w_{2m-1-i}$. We lift this restriction initially and assume a general window represented by a matrix D, a $2n \times 2n$ diagonal matrix. We can write $D = (D_0 \quad D_1)$ and D_r is a $n \times n$ diagonal matrix. Unlike for the LOTs, the matrix product $\Phi_0\Phi_0^*$ has no particular structure, in fact,

$$(\Phi_0\Phi_0^*)_{i,j} \;=\; a_{i,j} \;=\; \frac{1}{2m}\frac{\sin(\frac{\pi(i+j+1)}{2})}{\sin(\frac{\pi(i+j+1)}{2m})} + \frac{1}{2m}\frac{\sin(\frac{\pi(i-j)}{2})}{\sin(\frac{\pi(i-j)}{2m})}.$$

Substituting this into (4), we obtain the following:

$$a_{j,j}d_j^2 + (1-a_{j,j})d_{n+j}^2 = 1, \tag{16}$$

$$d_j d_s = d_{n+j}d_{n+s}, \quad s = 0, \cdots, n-1, s \neq j. \tag{17}$$

The set of solutions to (16)-(17) is infinite. Of course, the constant window with $d_j = 1$, for $j = 0, \ldots, 2n-1$ is also a solution to the above. If the window is symmetric, then (4) becomes:

$$D_0\Phi_0\Phi_0^*D_0 + JD_0J\Phi_1\Phi_1^*JD_0J \;=\; I \tag{18}$$

$$\text{with} \quad \Phi_0\Phi_0^* + \Phi_1\Phi_1^* \;=\; I. \tag{19}$$

Using (18), we derive the following conditions on D:

$$a_{j,j}d_j^2 + (1-a_{j,j})d_{n-j-1}^2 = 1, \tag{20}$$

$$d_j d_s = d_{n-j-1}d_{n-s-1}, \quad s = 0, \cdots, n-1, s \neq j. \tag{21}$$

Fixing $d_0 = -1$, we have $d_{n-1} = \pm 1$ and $d_s = -d_{n-1}d_{n-s-1}$ for $s = 1, \ldots, n-2$. Note that the same conditions hold for an antisymmetric window, that is, the half-windows can only be symmetric or antisymmetric. For a symmetric window, a possible solution is

$$d_j = \begin{cases} \cos(\frac{j\pi}{n-1} + \pi) & \text{if } n \text{ is even,} \\ \cos(\frac{2j\pi}{n-1} + \pi) & \text{if } n \text{ is odd,} \end{cases} \qquad j = 0, \ldots, n-1.$$

One of our first tasks in future work will be to use these symmetric windows in the PJB-LTFTs and study their effect. We also need to investigate window design techniques to obtain optimized windows that will modulate the PJB-LTFT and hence lead to better localization in the frequency band of the frame vectors.

5. REFERENCES

[1] K. Huang and R. F. Murphy, "Boosting accuracy of automated classification of fluorescence microscope images for location proteomics," *BMC Bioinformatics*, vol. 5, no. 78, 2004.

[2] Z. Kam, J. S. Minden, D. A. Agard, J. W. Sedat, and M. Leptin, "Drosophila gastrulation: Analysis of cell shape changes in living embryos by three-dimensional fluorescence microscopy," *Development*, vol. 112, pp. 365–370, 1991.

[3] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1995.

[4] H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Norwood, MA, 1992.

[5] J. Princen, A. Johnson, and A. Bradley, "Subband transform coding using filter bank designs based on time domain aliasing cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Dallas, TX, Apr. 1987, pp. 2161–2164.

[6] P. Hennings Yeomans, J. Thornton, J. Kovačević, and B. V. K. V. Kumar, "Wavelet packet correlation methods in biometrics," *Appl. Opt., sp. iss. Biometric Recogn. Systems*, vol. 44, no. 5, pp. 637–646, Feb. 2005.

[7] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part I)," *IEEE Signal Proc. Mag.*, vol. 24, no. 4, pp. 86–104, Jul. 2007.

[8] Z. Cvetković and M. Vetterli, "Oversampled filter banks," *IEEE Trans. Signal Proc.*, vol. 46, no. 5, pp. 1245–1255, May 1998.

[9] M. Püschel and J. Kovačević, "Real, tight frames with maximal robustness to erasures," in *Proc. Data Compr. Conf.*, Snowbird, UT, Mar. 2005, pp. 63–72.

[10] A. Chebira, Y. Barbotin, C. Jackson, T. Merryman, G. Srinivasa, R. F. Murphy, and J. Kovačević, "A multiresolution approach to automated classification of protein subcellular location images," *BMC Bioinformatics*, vol. 8, no. 210, 2007.

[11] R. A. Kellogg, A. Chebira, A. Goyal, P. A. Cuadra, S. F. Zappe, J. S. Minden, and J. Kovačević, "Towards an image analysis toolbox for high-throughput Drosophila embryo RNAi screens," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Arlington, VA, Apr. 2007, pp. 288–291.

[12] J. A. Ozolek, C. A. Castro, W. G. Jenkinson, A. Chebira, J. Kovačević, C. S. Navara, M. Sukhwani, K. E. Orwig, A. Ben-Yehudah, and G. Schatten, "Semiquantitative and multiresolution-based histological analysis of germ layer components in teratomas derived from human, non-human primate and mouse embryonic stem cells," Cairns, Australia, Jun. 2007.

[13] A. Chebira, L. P. Coelho, A. Sandryhaila, S. Lin, G. W. Jenkinson, J. MacSleyne, C. Hoffman, P. Cuadra, C. Jackson, M. Püschel, and J. Kovačević, "An adaptive multiresolution approach to fingerprint recognition," in *Proc. IEEE Int. Conf. Image Proc.*, San Antonio, TX, Sep. 2007, vol. 1, pp. 457 – 460.

[14] A. Chebira, J. A. Ozolek, C. A. Castro, W. G. Jenkinson, M. Gore, R. Bhagavatula, I. Khaimovich, S. E. Ormon, C. S. Navara, M. Sukhwani, K. E. Orwig, A. Ben-Yehudah, G. Schatten, G. K. Rohde, and J. Kovačević, "Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, May 2008, Submitted.

[15] A. Chebira and J. Kovačević, "Lapped tight frame transforms," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Honolulu, HI, Apr. 2007, vol. III, pp. 857–860.