

A MULTIREOLUTION ENHANCEMENT TO GENERIC CLASSIFIERS OF SUBCELLULAR PROTEIN LOCATION IMAGES

Thomas Merryman¹, Keridon Williams³, Gowri Srinivasa², Amina Chebira² and Jelena Kovačević^{2,1}

¹Electrical and Computer Engineering and ²Biomedical Engineering
Carnegie Mellon University, Pittsburgh, PA
³ Dept. of Science and Mathematics,
University of Virgin Islands, St Thomas, VI

ABSTRACT

We propose an algorithm for the classification of fluorescence microscopy images depicting the spatial distribution of proteins within the cell. The problem is at the forefront of the current trend in biology towards understanding the role and function of all proteins. The importance of protein subcellular location was pointed out by Murphy, whose group produced the first automated system for classification of images depicting these locations, based on diverse feature sets and combinations of classifiers. With the addition of the simplest multiresolution features, the same group obtained the highest reported accuracy of 91.5% for the denoised 2D HeLa data set. Here, we aim to improve upon that system by adding the true power of multiresolution—adaptivity. In the process, we build a system able to work with any feature sets and any classifiers, which we denote as a Generic Classification System (GCS). Our system consists of multiresolution (MR) decomposition in the front, followed by feature computation and classification in each subband, yielding local decisions. This is followed by the crucial step of combining all those local decisions into a global one, while at the same time ensuring that the resulting system does no worse than a no-decomposition one. On a nondenoised data set and a much smaller number of features (a combination of texture and Zernicke moment features) and a neural network classifier, we obtain a high accuracy of 89.8%, effectively proving that the space-frequency localized information in the subbands adds to the discriminative power of the system.

1. LOCATION PROTEOMICS AND MULTIREOLUTION ANALYSIS

Among the most important goals in biological sciences today is to understand the role and function of all proteins. One of the critical aspects of a protein's activity and function is its subcellular location (PSL), that is, the spatial distribution of proteins within the cell. Today's method of choice to determine PSL is fluorescence microscopy, its success due in part to the advent of a range of new fluorescent probes used to tag proteins or molecules of interest, including the nontoxic, green fluorescent protein (GFP). Once successfully tagged, the cells are imaged using one of many models of fluorescence microscopes to produce a multidimensional data set—a *bioimage*. These bioimages can be just 2D slices or 3D cell/tissue volumes (z-stacks). Acquiring bioimages at multiple time instants results in 3D movies (2D time series) or 4D data sets (z-stacks tracked in time). Finally, these microscopes allow for imaging of multiple fluorescence channels, bringing the possible dimensionality of the

This work was supported in part by the PA State Tobacco Settlement, Kamlet-Smith Bioinformatics Grant.



Fig. 1. MR enhancement to a generic classification system.

data set to 5D. With the enormous volume of such high-dimensional data sets being generated, human analysis becomes time-consuming, prone to error and ultimately, impractical, leading to the “holy grail” for PSL bioimage interpretation and analysis: develop a system for *fast*, *automatic*, and *accurate* recognition of proteins based on their subcellular location images.

Murphy et al. pioneered automated PSL interpretation and analysis, resulting in a system that can classify protein location patterns with well-characterized reliability and better sensitivity than human observers [1, 2]. This work was followed by [3, 4]. With the addition of the simplest multiresolution features, in [1], the authors obtained the highest reported accuracy of 91.5% for the denoised 2D HeLa data set.

Problem Statement. The problem we are addressing is that of classifying the spatial distribution patterns of selected proteins within the cell. The challenge in this data set is that images from the same class look different while those from different classes look very similar (see Fig. 2). In the data sets we use, the proteins were labeled using immunofluorescence. So we know the ground truth, that is, which proteins were labeled and subsequently imaged. This is useful for algorithm development as we can test the accuracy of our classification scheme.

Methodology. As the introduction of the simplest multiresolution features produced a statistically significant jump in classification accuracy, our aim is to explore more sophisticated multiresolution techniques. In particular, the following are the three characteristics of multiresolution we wish to explore:

(a) Localization: Fluorescence microscopy images have highly localized features both in space and frequency. This leads us to MR tools, as they have been found to be the most appropriate tools for computing and isolating such localized features [5].

(b) Adaptivity: Given that we are designing a system to distinguish between classes of proteins, it is clear that an ideal solution is adaptive, a property provided by MR techniques.

(c) Fast and Efficient Computation: It is well known that wavelets have a computational cost of the order $O(n)$, where n is the input size, as opposed to $O(n \log n)$ typical for other linear transforms including the FFT.

Our Philosophy. Based on the above arguments, we would like to extract discriminative features within space-frequency localized subspaces. These are obtained by MR decomposition; thus, instead

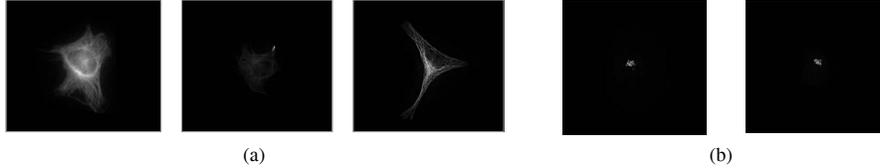


Fig. 2. (a) Intra-class variation: The three images show the spatial distribution of Tubulin within a cell. (b) Interclass similarity: The first image shows the spatial distribution of Giantin and the second image shows the spatial distribution of Gpp130. Both are Golgi-proteins.

of adding MR features as in [1], we compute features in the MR-decomposed subspaces. Thus, our system is a generic system with an MR decomposition block in front (see Fig. 1), so there is feature computation and classification in each of the subspaces. These are combined through a weighting process.

2. BACKGROUND AND PREVIOUS WORK

Automated System for Location Classification. The heart of the previously-developed system is a set of numerical features—termed Subcellular Location Features (SLF)—to describe the spatial distribution of proteins (protein location patterns) in each cell image. The SLFs are drawn from various categories of image analysis methods, including texture analysis, decomposition methods, and morphological image processing. In [1], the previous system was enhanced with the simplest, off-the-shelf, multiresolution (MR) features—wavelet and Gabor. From this feature set, a discriminating subset (SLF15) was chosen using stepwise discriminant analysis. The results were very encouraging: With a mixture of expert classifiers, this set can produce an average accuracy of 91.5% on single-cell images. This is compared to the previous best result of 89.4%, a statistically significant jump. This improved performance with the addition of the simplest, standard MR features, was the impetus behind our hypothesis that adding the true power of MR— adaptivity, should result in improved performance.

MR Techniques. We argued in Section 1 that the nature of our data sets requires tools which offer localization in space and frequency as well as adaptivity, and we further argued that those tools are MR in nature. MR transforms are many; we now give a brief overview. The basic idea behind MR is that we can look at a signal at different scales or resolutions and at different points in time. This should give us information about hidden structures in the signal, with a particular behavior across scales.

The main building block of any MR transform is a *filter bank* [6]; it is a device that splits the signal into *subbands*, where each subband contains one part of the signal’s spectrum. In Fig. 3, a 2D filter bank is given. On both rows and columns, the signal is filtered, followed by downsampling (discarding every other sample, allowed because there is filtering beforehand). In the simplest case, this produces 4 subbands; one extracting lowpass information in both directions, one extracting highpass information in both direction and the remaining two extracting lowpass information in one direction and the highpass information in the other.

Adaptivity of MR transforms manifests itself differently; many popular transforms are subsumed in this scenario, such as: (a) Growing a full tree to L levels with filters the same length as the downsampling factor yields the Discrete Fourier Transform (DFT) of size 2^L . (b) Growing a full tree to L levels but allowing the filters to be longer, leads to the Short-Time Fourier Transform, or, the Gabor Transform. (c) Growing the tree only on the lowpass branch to

| | Average Accuracy [%] | | |
|-----------|----------------------|--------------|---------------------------|
| | Morphological only | Texture only | Morphological and Texture |
| No MR | 53.6 | 71.8 | 69.4 |
| MR | 68.9 | 82.2 | 80.2 |

Table 1. The average accuracy per class from [8]. If no multiresolution is used, the texture features outperform morphological, and together they achieve the accuracy of 69.4%, lower than texture features only. Adding MR improves the accuracy for each set and yields the accuracy for the combined features of 80.2%, lower then for texture features only.

L levels leads to the Discrete Wavelet Transform (DWT) of L levels. (d) Growing an arbitrary tree leads to Wavelet Packets (WP). (e) Splitting the signal into more than two channels, allowing filters in the above transforms to be orthogonal and/or linear phase, etc., leads to even more degrees of freedom.

The wavelet packets mentioned above [7], adapt themselves to the signal at hand. However, this is possible only if a suitable cost function is available. Given that we had no natural cost function available, we decided to first test whether the space-frequency adaptivity offered by the MR transform help, and, if it does, embark on the search for a suitable cost function.

Adaptive MR Maximum-Likelihood Classifier. Our first step was the algorithm developed in [8]; we summarize it here briefly. We started with a simple classification system consisting of Haralick texture feature computation followed by a maximum-likelihood classifier, and demonstrated that, by adding an MR block in front, we are able to raise the average classification accuracy by roughly 10% (see Table 1). This fits within our generic framework as in Fig. 1, where the feature computation block uses Haralick texture features and the classification block is maximum likelihood. We concluded that selecting features in MR subspaces allows us to custom-build discriminative feature sets for fluorescence microscopy images of protein subcellular locations. We then added morphological features and found that, while MR increases accuracy considerably in all cases, our classifier is probably not powerful enough as combining morphological and texture features did not raise the accuracy, and even lowered it by 2%. Our aim is to improve upon this result.

3. PROPOSED ALGORITHM

As we just discussed, although in [8] we demonstrated the power of adaptive MR techniques, problems still remain, mostly due to the classifier we used (maximum likelihood). The algorithm adjusts the

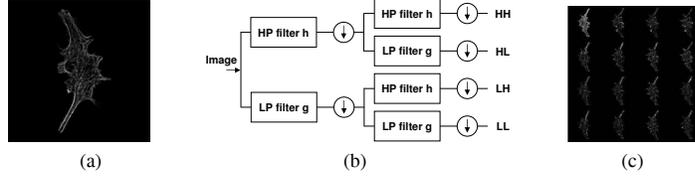


Fig. 3. (a) An image of Actin. (b) Two channel analysis filter bank for 2D signals. The filter h is a highpass filter and g is a lowpass filter. (c) A 2-level MR filter bank decomposition of Actin. The upper left image is the coarse representation. The other images are the detailed representations.

weights trying to maximize the overall classification accuracy; however, the previous best is not kept. We want to try to rectify that. Moreover, we would like to incorporate MR techniques to enhance a generic classifier. We define a generic classifier as one that takes as input a vector of features, and outputs a decision vector that matches the length of the number of classes.

There is no reason why our approach in [8] would not work if we viewed each of the subbands as an individual classification branch and applied a generic classifier to it. This would enable us to come up with individual classification decisions at each particular subband. What we propose to do is consolidate all of the decomposed subband decisions into one decision we believe will be of higher accuracy than each of the individual decisions. We aim to achieve this result by using an adaptive weighting algorithm.

Fig. 4 shows a graphical representation of the process of concatenating all of the subband decisions into one. We use weights at each subband to weigh 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 will without the adaptive MR system in place. With this information, we know that there is a weight combination that will do at least as well as the generic classifier. Our goal is to decide how to find the weight vector that achieves the highest overall classification accuracy to the data set. The algorithms for both the training and the testing phases are given in Algorithms 1 and 2.

4. EXPERIMENTAL RESULTS

Data Sets. We used the 2D HeLa collection of cervical cancer cells

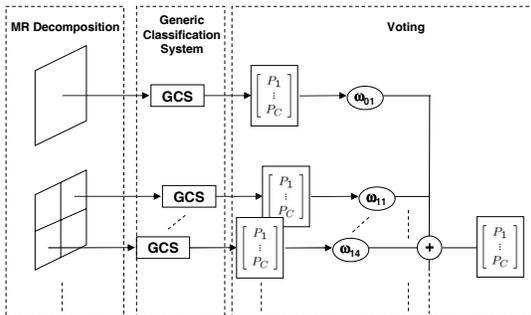


Fig. 4. MR enhancement to generic classifiers. The image is first decomposed using a level- L filter bank. Classification is ran on each of the subbands leading to a probability vector which are then weighed and combined to give a final probability vector.

Algorithm 1 [Training Phase] Input: D (local decision vectors). Output: W (weight vector).

TrainingPhase(D)

```

initialize  $w_{1,s}$  to classification accuracy of subband ( $D_s$ )
initialize  $w_{2,s}$  to emphasize the 0th subband
initialize  $w_{3,s}$  to positive random entries
for all weight vectors,  $a = 1$  to 3 do
  normalize  $w_a$ , initialize counter,  $c = 0$ 
  compute classification accuracy with  $w_a$  and store in  $p_a$ 
  while  $c < \text{maxEpochs}$  do
    increment counter,  $c = c + 1$ 
    for all images,  $n = 1$  to  $N$  do
      set  $d =$  the image is classified correctly with  $w_a$ 
      if  $d$  is incorrect then
        for all subbands,  $s = 0$  to  $S - 1$  do
          set  $t =$  subband classified correctly
          if  $t$  is correct then
             $w_{a,s} = w_{a,s} \cdot (1 + \epsilon)$ 
          else
             $w_{a,s} = w_{a,s} \cdot (1 - \epsilon)$ 
          end if
        end for
      end if
    end for
    compute classification accuracy with  $w_a$  and store in  $p_a^{new}$ 
    if  $p_a^{new} > p_a$  then
      set  $p_a = p_a^{new}$ , save  $w_a$  as  $w_a^{best}$ , reset counter,  $c = 0$ 
    end if
  end while
end for
set  $W$  to the  $w_a^{best}$  with the greatest  $p_a$ 
return  $W$ 

```

Algorithm 2 [Testing Phase] Input: D (local decision vectors), W (weight vector). Output: G (global decision vector), A (classification accuracy).

TestingPhase(D, W)

```

set  $G = \sum_{s=0}^{S-1} D_s \cdot W_s$ 
set  $A$  equal to the classification accuracy of  $G$ 
return  $G$  and  $A$ 

```

with 50 single-cell images of size 512×512 , in each class, courtesy of the MurphyLab [9]. There were 10 classes of subcellular location patterns obtained by marking the endoplasmic reticular protein, Giantin and Gpp130, two golgi proteins, a lysosomal protein, a mitochondrial protein, Nucleolin (nucleolar protein) Actin, Tfr (endoso-

mal protein) Tubulin and the DNA was also labeled. It is important to note that this data set was not put through some of the pre-processing steps such as background thresholding and rotation of the images to a common reference point, while the results reported in [1] were on the pre-processed data set. In future work, we will try to access the the original set and intend to conduct experiments that will allow us to compare to our algorithms. For now we can only speculate on the effects of the preprocessing. One hypothesis is that since the data is denoised, it is easier to classify as the classifier does not get confused by extra noise. On the other hand, it is possible that the pre-processing removes parts important for classification and thus makes it harder to classify. In our future work, we also hope to address the problems associated with translation and rotation sensitivity of our classification system.

MR Decomposition. We decompose our images using a simple, separable, 2D filter bank with Haar filters (see Fig. 3), up to 2 levels, yielding a total of 21 subbands, corresponding to all the nodes in the decomposition (1 node at level 0, 4 nodes at level 1 and 16 nodes at level 2, for a total of 21).

Feature Computation. As in [1], we use Haralick texture features (13), morphological (16) and Zernicke moments (49). Unlike in [1], we do not use wavelet/Gabor features as the MR advantage brought by those has been taken over by our MR decomposition. Therefore, our total number of features is 78, as opposed to 180 in [1]. Instead of combining all features in a single probability vector, we allow each feature set its own probability vector per subband; effectively, this brings the number of subbands to $3 \cdot 21 = 63$. Note that while we have decreased the number of features significantly, we have increased the number of classifiers, as now we have a classifier per subband. Understanding the computational trade-off is a task for future work.

Classification. We used a two-layer neural network 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. With this layout, there is an input for each feature and an output for each class. We then train the neural network using a one-hot design. Since each output from the second layer corresponds to a class, when training, 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. To maximize the use of our data, our training process uses five-fold cross validation. This ensures that we have an ample amount of training data without sacrificing the use of all of the images in our adaptive weighting phase. We train for 25 epochs. We also perform ten-fold cross validation on the weight calculation.

Results. We ran our algorithm with and without the MR block, for each feature set, for combinations of two feature sets, as well as all feature sets together. There were two schemes to weighting: (1) As we explained above, we have 21 subbands/feature set, giving a total of 63 subbands which are then weighted all together. (2) Perform weighting on each individual feature set and then weight the resulting three numbers. As the second scheme gives slightly higher accuracies, this is the one we will be using.

We see that, without fail, MR block increases the classification accuracy in each case; the increase is the largest for single feature sets (Z and M). This brings another trade-off into the game; instead of increasing the accuracy by adding more features, increase it by adding the MR block. It is worth noting that the best result of almost 90% is obtained by combining texture and Zernicke features; adding morphological features reduces accuracy slightly (for MR).

We conclude that with a much smaller number of features than in [1], we obtain an accuracy of close to 90%. This number may

| | Average Accuracy [%] | | | | | | All |
|-----------|----------------------|------|------|------|------|-------------|------|
| | Z | M | T | M,Z | T,M | T,Z | |
| No MR | 51.4 | 69.6 | 86.6 | 74.8 | 86.6 | 86.6 | 87.4 |
| MR | 62.8 | 81.0 | 89.2 | 83.8 | 89.2 | 89.8 | 89.2 |

Table 2. The average accuracy per class. Z, M and T stand for Zernicke, morphological and texture features.

change when we gain access to the exact same data set as in [1]; that set was denoised while ours was not. There are a number of other issues to be explored: For example, our system effectively builds an adapted MR decomposition (via subband weights) for the whole data set; we need to adapt that decomposition to each class. We also need to test the robustness of the system to other data sets. Computational burden is also an issue; we will address it in an upcoming paper.

In summary, to explore space-frequency localized information in our data, we have built a flexible add-on to a generic classification system in the form of MR decomposition and voting. The system performs well and adds to the accuracy of a nondecomposed system. **Acknowledgments.** We gratefully acknowledge the help and support from MurphyLab at Carnegie Mellon University, in particular, Ms. Elvira Garcia Osuna and Prof. Robert F. Murphy.

5. REFERENCES

- [1] K. Huang and R. F. Murphy, "From quantitative microscopy to automated image understanding," *Journ. Biomed. Optics*, vol. 9, pp. 893–912, 2004.
- [2] R. F. Murphy, "Location proteomics: A systems approach to subcellular location," *Biochem. Soc. Trans.*, vol. 33, pp. 535–538, 2005.
- [3] P. Perner, H. Perner, and B. Muller, "Mining knowledge for Hep-2 cell image classification," *Journ. Artificial Intelligence in Medicine*, vol. 26, pp. 161–173, 2002.
- [4] A. Danckaert, E. Gonzalez-Couto, L. Bollondi, N. Thompson, and B. Hayes, "Automated recognition of intracellular organelles in confocal microscope images," *Traffic*, vol. 3, pp. 66–73, 2002.
- [5] S. Mallat, "Wavelets for a vision," *Proc. IEEE*, vol. 33, pp. 604–614, 1996.
- [6] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [7] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," Tech. Rep., Yale University, 1991.
- [8] G. Srinivasa, T. Merryman, A. Chebira, A. Mintos, and J. Kovačević, "Adaptive multiresolution techniques for subcellular protein location image classification," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Toulouse, France, May 2006, Invited paper.
- [9] The Murphy Lab at Carnegie Mellon University: <http://murphylab.web.cmu.edu>, 1998.