Features for Protein Localization in Microscopic Cellular Images

Song Liu, Jagath C Rajapakse

Abstract—**Localization of proteins of microscopic images is an important problem in bioimaging. In this manuscript, we propose three kinds of novel features of microscopic images: steerable filter feature, tree-structured wavelet transform feature, and gray scale and invariant texture features. We show that these features are useful in improving the accuracy of classification of protein localization of proteins in specific subcellular locations.**

***Index Terms*— bioimaging, feature extraction, microscopic images, pattern recognition, protein subcellular localization.**

I. Introduction

An important aspect of identifying the function and characterization of a protein is the determination of its localization of subcellular organelles. An automated system for analyzing interpreting images of localization patterns were demonstrated to successfully distinguish up to ten subcellular patterns in Hela cells [1]. The features extracted from images play an important role in determining accuracy of classification accuracy of protein subcellular localization [2]. Because of the similarity between certain classes (such as endosomal and lysosomal proteins), it is still useful to improve the classification accuracy by incorporate additional features. The classification accuracy was improved from 83% with 84 features to 92% with 180 features [3]. Compared to the former feature set, we proposed new features that are useful in distinguishing similar proteins. In this paper, we explore the probability of enhancing the classification accuracy through new kinds of features — steerable filter feature[4], tree-structured wavelet transform feature [5] and gray scale and invariant texture features [6].

II. METHODS

The microscopic images of the Hela cells were downloaded from: http://murphylab.web.cmu.edu/data/ and were preprocessed as the procedure described by Boland [1]. Because our aim of this paper is to test the efficiency of the new features in classifying protein localization, we used the preprocessed images of three kinds of subcellular proteins (see Figure 1), namely, Endoplasmic Reticulum (ER, labeled with ER protein), Golgi_gia (labeled with giantin), and Endosome (labeled with TfR (Transferrin Receptor). All the images were collected using fluorescent microscopy with a 100×objective and 23-μm square pixels.

The Subcellular Location Features (SLF) defined by Murphy et al. [1] was derived by methods developed by their team. These features included 22 morphological and edge features, 49 Zernike features, 13 haralick texture features, 60 Gabor texture features and 30 Daubechies 4 wavelet features. At first, we chose zernike features to compare the classification accuracy with our newly developed features. Then the steerable filter features were extracted and these features described the maximum energy, minimum energy, max/min ratio, mean, standard deviation of the transformed images. Four tree-structured wavelets transform feature were obtained by determining the dominant channels of decomposing the images [7]. Finally, we computed 256 gray scale and rotation invariant features which are similar to the Local Binary Patterns developed by Timo Ojala [6].

After extraction of image features, these features were used to train a neural network (NN) classifier and test the classification accuracy. In our approach, we have chosen three kinds of subcellular structures and 30 samples for each. In fact, a fixed number of instances from each class will be randomly assigned to the train, stop, and test sets. For each epoch, 48 instances from different groups were used to train the network, 24 instances were used for stop data and 18 instances were used for test. The number of inputs equals to the number of features being evaluated. There are 20 hidden nodes, and 10 output nodes. The training epoch is 50 times to get the optimal setting. The target outputs of the network for each instance were defined to be 0.9 for the node representing the correct class of that instance and 0.1 for the other outputs. After each epoch of training, the stop data were passed through the network for calculating a sum of squared error between the actual network outputs and the target outputs. When this error term for the stop data reached a minimum, training was halted. Then the test data would be applied and the outputs recorded.

S Liu and J. C. Rajapakse* are with the BioInformatics Research Centre and the School of Computer Engineering, Nanyang Technological University, Singapore, (*corresponding author tel: +65 67905802; fax: +65 67926559; e-mail: asjagath@ ntu.edu.sg).

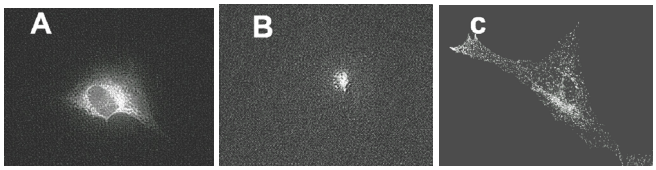

Fig.1. The fluorescent microscopic images of three subcellular proteins in HeLa cells: (A) Endoplasmic Reticulum, labeled with ER protein; (B) Golgi_gia labeled with giantin; and (C) Endosome labeled with TfR (Transferrin Receptor).

III. RESULTS

As seen from tables 1-4 for classification accuracy, the steerable filter features demonstrated good performance in classifying Giantin (the classification accuracy is 95.00%), compared to Zernike features which showed good result in specifying ER (96.67%),. The performance of these two groups of features was very similar in specifically classifying certain kinds of proteins. For the tree-structured wavelet transform features, the classification accuracy is comparatively good for ER and Giantin. We believe that this kind of feature would perform better when combining with other features. Finally, the gray-scale and rotation invariant features performed better in classifying all the three kinds of proteins.

Table 1. Performance of Zernike features for classifying three kinds of subcellular proteins.

True classification	Output of classifier		
	ER	Giantin	TfR
ER	0.9667	0	0.0333
Giantin	0.1333	0.5667	0.3000
TfR	0.3667	0.3167	0.3167

Table 2. Performance of Steerable filter features for classifying three kinds of subcellular proteins

True classification	Output of classifier		
	ER	Giantin	TfR
ER	0.3833	0.2000	0.4167
Giantin	0.0333	0.9500	0.0167
TfR	0.2167	0.3167	0.4667

Table 3. Performance of tree-structured wavelet features for classifying three kinds of subcellular proteins

True classification	Output of classifier		
	ER	Giantin	TfR
ER	0.7000	0.0667	0.2333
Giantin	0.0833	0.6667	0.2500
TfR	0.0500	0.7667	0.1833

Table 4. Performance of gray-scale and rotation invariant features for classifying three kinds of subcellular proteins.

True classification	Output of classifier		
	ER	Giantin	TfR
ER	0.8500	0.0333	0.1167
Giantin	0.0333	0.8333	0.1333
TfR	0.0500	0.1833	0.7667

IV. CONCLUSION

We have developed three kinds of features for classification of subcellular localization patterns from microscopic images: steerable filter features, tree-structured wavelet transform features and gray-scale and rotation invariant features. Tests on classification of three kinds of subcellular protein fluorescence images showed that Steerable filter features and tree-structured wavelet transform features would contribute to improvement in classifying location of certain subcellular proteins, namely, Endoplasmic Reticulum and Golgi_gia. The gray-scale and rotation invariant features showed the potential in improving the classification accuracy for all the three kinds of subcellular proteins considered in the experiments.

REFERENCES

[1] M. V. Boland, R. F. Murphy, "A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence images of HeLa cells," *BIOINFOMATICS,* vol. 17, No. 12, pp.1213-1223, August. 2001.

[2] K. Huang, R. F. Murphy, "Boosting accuracy of automated classification of fluorescence microscope images for location proteomics," *BMC Bioinformatics* 2004, 5:78 doi:10.1186/1471-2105-5-78.

[3] S. C. Chen, R. F. Murphy, "A graphical model approach to automated classification of protein subcellular location patterns in multi-cell images," *BMC Bioinformatics* 2006, 7:90 doi:10.1186/1471-2105-7-90.

[4] X. D. Yang, H. Q. Li, X.B. Zhou, S. T. C. Wong**, "**Identification of cell-cycle phases using neural network and steerable filter features**,"** Lecture Notes in Computer Science, Advances in Neural Networks, vol. 3973, pp.702 – 709, 2006.

[5] T. H. Chang, C.-C. Jay Kuo, **"**Texture analysis and classification with tree-structured wavelet transform," *IEEE Transactions On Image Processing*, vol. 2, No. 4, pp. 429-441, Octobor. 1993.

[6] T. Ojala, M. PietikaÈäinen, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 24, No. 7, pp.971-987, July. 2002.

[7] N. Harder, F. Mora-Bermúdez, W.J. Godinez, J. Ellenberg, R. Eils, K. Rohr, "Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences," *Proc. MICCAI'06*, vol. 4190 of *LNCS*, Copenhagen, DK, pp. 840-848, 2006. Springer-Verlag.