DOI: 10.1017/S1431927604886744
Microsc Microanal 10(Suppl 2), 2004

Automated Determination of Subcellular Location from Confocal Microscope Images

R. F. Murphy*

* Departments of Biological Sciences and Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

Confocal microscopy is frequently used to examine the subcellular distributions of proteins or other biological macromolecules, but essentially all interpretation of the patterns in the resulting images is done by visual inspection. The difficulties associated with this approach include lack of reproducibility of assignments (both from viewer to viewer and from image to image), lack of precision in assigning location (especially for mixed patterns or patterns constituting a subdomain of an organelle, and potential inability to distinguish subtle changes in populations of cells. Automating this process allows it to be done more objectively, more rapidly, and more accurately.

With this in mind, my group has previously described systems for automated recognition of the major subcellular structures in 2D fluorescence microscope images of cultured cells [1-3]. These systems are based on sets of numerical features (which we term Subcellular Location Features, or SLFs) that capture the essence of the subcellular pattern without being overly sensitive to the size, shape, position and orientation of each cell. We have described 180 features of a variety of types, including 6 features that describe a protein pattern with reference to a parallel image of total DNA [4]. Feature selection by Stepwise Discriminant Analysis revealed that a set of 47 of these features in combination with a mixture-of-experts classifier can achieve over 92% accuracy for classifying ten subcellular patterns in 2D images of HeLa cells. Of particular importance was the observation that two Golgi proteins included in the set could be distinguished with over 86% accuracy even though they could not be distinguished beyond random accuracy by visual examination [3].

The success obtained in classifying subcellular patterns in 2D images led us to investigate whether improvement could be obtained by acquiring 3D images and implementing features to take advantage of the additional information in them [5]. To enable fully automated analysis of images that might contain more than one cell, we acquired three-color confocal images using probes for total DNA and total protein in addition to a probe for each of nine different proteins (an example is shown in Fig. 1). The total DNA and total protein images were used to create single-cell regions by a seeded watershed algorithm. Classification accuracy for the 3D images improved to 96% (with over 97% accuracy in distinguishing the two Golgi proteins) [4]. The improvement in accuracy over 2D images was in part due to the fact that 2D images do not always capture the most informative region of a cell and in part due to the additional information present in 3D images.

The finding that subcellular patterns can be recognized with sensitivity and specificity higher than human observation using SLFs validates the use of SLFs for comparing protein patterns (e.g., in the presence and absence of a drug) and for clustering proteins by their subcellular patterns. They also indicate the value of collecting full 3D images when determining protein subcellular location [6].

References

[1] M.V. Boland et al., *Cytometry 33:* (1998) 366.

[2] M.V. Boland and R.F. Murphy, *Bioinformatics 17:* (2001) 1213.
[3] R.F. Murphy et al., *J VLSI Sig Proc 35:* (2003) 311.
[4] K. Huang and R.F. Murphy, *submitted:* (2004).
[5] M. Velliste and R.F. Murphy, *2002 IEEE International Symposium on Biomedical Imaging (ISBI-2002)*. (2002) 867.
[6] This research was supported in part by NSF grants BIR-9217091, MCB-8920118, and BIR-9256343, by NIH grants T32 GM08208 and R33 CA83219, and by a research grant from the Commonwealth of Pennsylvania Tobacco Settlement Fund. 3D imaging of HeLa cells was made possible by the generous assistance of Dr. Simon Watkins.

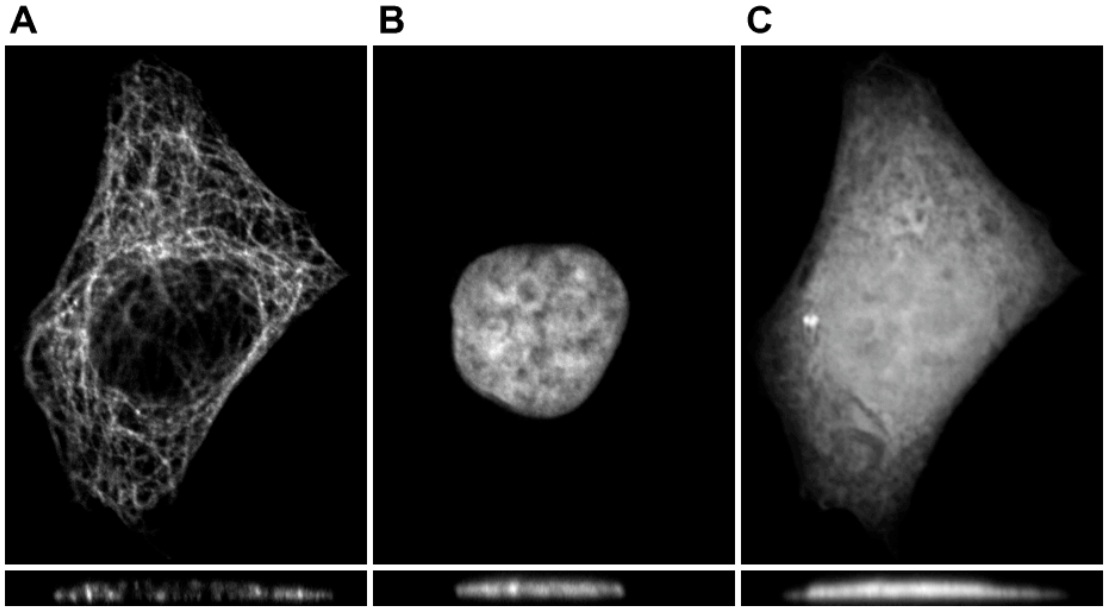

FIG. 1. An example image from the 3D HeLa data set collected with a confocal microscope. Each image in the data set consisted of three channels corresponding to a specific protein label (A) (tubulin in this example), a DNA label (B), and a total protein label (C). The images shown represent horizontal (top) and vertical (bottom) slices through the middle of the cell, chosen from the full 3D images to intersect the center of fluorescence of the specific-protein channel. (From reference [5].)

TABLE 1. Confusion matrix for automated recognition of subcellular patterns in 3D HeLa cell images. The average accuracy was 95.8%. (From reference [4].)

	Cyt	DNA	ER	Gia	Gpp	Lam	Mit	Nuc	Act	TfR	Tub
Cyt	100	0	0	0	0	0	0	0	0	0	0
DNA	0	98.1	0	0	0	0	0	1.9	0	0	0
ER	0	0	96.6	0	0	0	0	0	1.7	0	1.7
Gia	0	0	0	98.2	0	1.9	0	0	0	0	0
Gpp	0	0	0	4	96.0	0	0	0	0	0	0
Lam	0	0	0	1.8	1.8	96.4	0	0	0	0	0
Mit	0	0	0	3.5	0	0	94.7	0	1.8	0	0
Nuc	0	0	0	0	0	0	0	100	0	0	0
Act	0	0	1.7	0	0	0	1.7	0	94.8	1.7	0
TfR	0	0	0	0	0	5.7	3.8	0	1.9	84.9	3.8
Tub	0	0	3.7	0	0	0	0	0	0	1.9	94.4