A Distributed Database for Bio-Molecular Images

Ambuj K. Singh
Department of Computer Science
University of California
Santa Barbara, CA 93106
ambuj@cs.ucsb.edu

B. S. Manjunath
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106
manj@ece.ucsb.edu

Robert F. Murphy
Departments of Biological Sciences and Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
murphy@cmu.edu

ABSTRACT

Information technology research has played a significant role in the genomics revolution over the past decade, from aiding with large-scale sequence assembly to automating gene identification to efficiently searching databases by sequence similarity. The tremendous amount of information gathered from genomics will be dwarfed in the next decade by the knowledge to be gained from comprehensive, systematic studies of the properties and behaviors of all proteins and other biomolecules. High-resolution imaging of molecules and cells will be critical for understanding complex systems such as the nervous system, whether it be for the localization of specific neuron types within a region of the central nervous system, the branching pattern of dendritic trees, or the localization of molecules at the subcellular level. Furthermore, knowing how these distribution patterns and subcellular locations change as a function of time is critical to understanding how cells respond to stress, injury, aging, and disease. We are developing sophisticated information technologies for collecting and interpreting the enormous volume of biological image data. A major outcome of the research will be a unique, fully operational, distributed digital library of biomolecular image data accessible to researchers around the world. Such searchable databases will make it possible to optimally understand and interpret the data, leading to a more complete and integrated understanding of cellular structure, function and regulation.

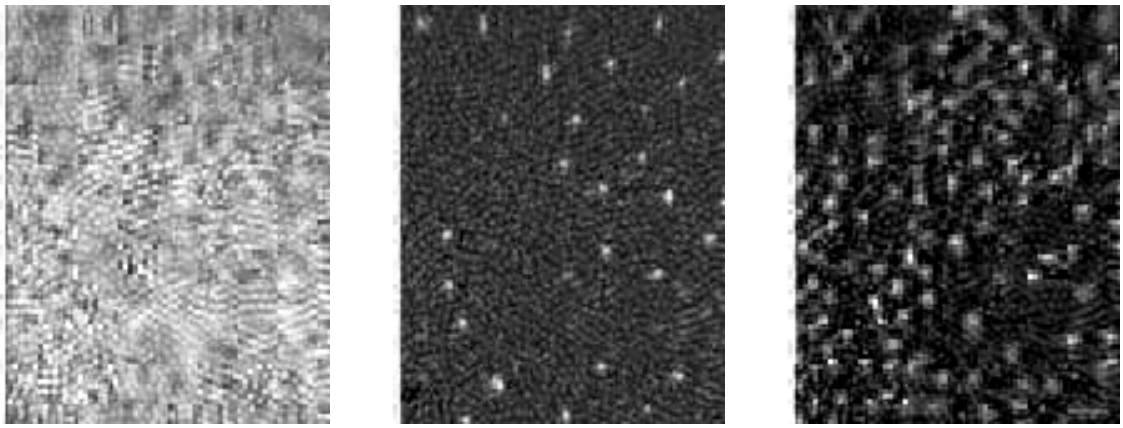
1. INTRODUCTION

Significant progress in our understanding of cellular and sub-cellular events can be made if we can couple advances in information technologies, such as image processing, pattern recognition, and databases, with the enormous volume of biomolecular images that are being generated. For example, biologists make extensive use of fluorescence microscopy to achieve high sensitivity and subcellular resolution. One strategy makes use of fluorescently labeled antibody probes to visualize almost any molecule(s) of interest in any cell or tissue that has been tagged. Viewing these samples with a confocal microscope yields thin optical sections of each sample. Generation of a z-series composed of numerous sequential sections allows precise reconstruction of the entire cell. The resulting data are often complex and full of subtleties. The data become even more complex when imaging multiple proteins through multiple, independent channels. Two to four channels are common, but this number will increase with advances in multispectral imaging and fluorescent dye technology.

A second strategy developed in the past decade makes use of fusions of green fluorescent protein (GFP) to proteins of interest. It has made it possible to simultaneously visualize multiple molecules of interest in living cells in real time, providing unprecedented insights into the *in vivo* action of these proteins and the relationships among them. In addition, atomic force microscopy (AFM) has recently become an important tool for imaging of biological molecules. With AFM, a sample is analyzed by probing the surface with a tip, and the interaction between tip and sample is measured. Physical topography, charge density, magnetic field, temperature and other surface properties can be discerned. Given its high resolution and multi-dimensional capabilities, its application to biological issues is certain to increase dramatically, carrying with it the generation of voluminous data requiring precise analysis.

Each of the strategies described above is applicable to an enormous number of molecular and cell biology questions and investigations. Antibody probes specific to thousands of different proteins are available commercially. In addition to identifying particular molecules, a subset of these antibody probes recognize particular conformational (i.e., functional) molecular states [13, 26]. With respect to the GFP approach, recombinant DNA technology makes it straightforward to fuse GFP to any protein of interest [18, 21, 29]. The immense increase in the number of antibody probes that can be used for immunolocalization, along with the ease of image capture by techniques such as laser-scanning con-

SIGMOD Record



(a) red (b) green (c) blue

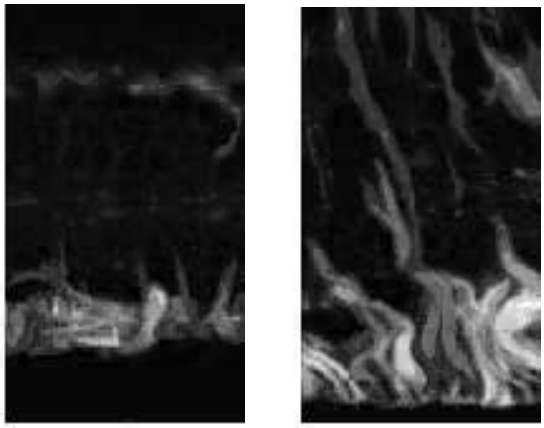
Figure 1: An array of photoreceptors labeled with specific antibodies to the long-wavelength cones (red), short-wavelength cones (green), and rods (blue). The three channels are normally combined in one psudocolored image, but are separated here.

focal microscopy and atomic force microscopy have resulted in an explosion of the amount of biological information available in the form of digital images. However, there is currently no central home for this vast amount of data, and no method readily available to discover knowledge in such a database were it available. The primary goal of this project is to develop new information processing technologies that enable the scientific community to take full advantage of the knowledge embedded in these large data sets.

2. CENTRAL NERVOUS SYSTEM

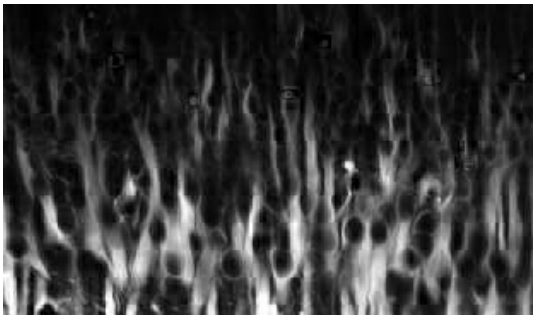
The central nervous system (CNS) is a major focus of the project. Our goal is to provide tools that may help unravel the functional secrets of this immensely complex system. In the following, we give some examples of database problems that, if successfully integrated with the appropriate smart imaging and information processing technologies, will radically advance our understanding of some of the fundamental processes at the cellular and subcellular level.

The vertebrate retina has been widely studied at the cellular level for over 150 years. It is the initial site of both optical image formation and the neural processing that leads to the formation of an image within the higher visual centers. Because different neurons express different sets of proteins, different antibodies can be used as probes to examine the distribution pattern of cells within a tissue [7] The example in Figure 1 shows the array of photoreceptors labeled with specific molecular probes to the short-wavelength cone, long-wavelength cone, and rod photoreceptors. This array of receptor types varies with retinal region as well as across species. It will change during development, and in response to trauma or an inherited disease. Understanding if the patterns are maintained across species, how and when they emerge in development, and whether or not they change after injury, during specific diseases, or during

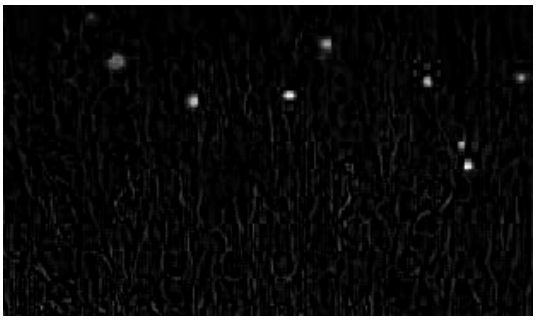


(a) normal (b) injured

Figure 2: An example of the distribution of a filamentous protein in (a) a normal retina and (b) a detached retina. The protein's pattern of localization shifts dramatically as a result of the injury.



(a) green



(b) red

Figure 3: A retinal section labeled with an antibody to an intermediate filament protein (green) and a probe that recognizes dying cells (red).

aging is a problem of immense importance but also of immense complexity.

Proteins can be localized in many different patterns within a specific cell. But the exact pattern of intracellular localization of a "vesicular" or "filamentous" pattern may vary tremendously depending upon the molecule being studied, the cell type, or the developmental or metabolic state of the cell. Figure 2 shows an example of a filamentous protein that is localized to a specific domain in normal cells. The amount of the protein and the pattern of its localization shift dramatically when the retina is injured or in response to specific disease states [23, 24]. Populations of different vesicles, each carrying a unique set of protein molecules will shift in localization depending upon, for example, their target. Some may remain within the cell body, others may be transported specifically into dendrites, while still others may be transported into axons. Thus, the subcellular pattern of localization of these proteins is fundamental to unraveling their functional significance.

During development of the nervous system, many more neurons are produced than are ultimately used. These "overproduced" neurons die by programmed cell death (apoptosis), and this cell death is thought to occur by specific spatial or temporal patterns within a given tissue or region of the CNS. Cell death also occurs during specific diseases. There are now over 100 gene mutations in several different photoreceptor molecules that each lead to the disease phenotype known as "retinitis pigmentosa." In all cases, photoreceptor cells die by apoptosis. There are specific markers that can be used to image apoptotic cells (Figure 3). Knowing the spatial pattern of cell death along a temporal dimension for each of these mutations may reveal the functionality of the gene product or the mechanism resulting in cell death. For example, do photoreceptor cells die as a "wave" across their specific layer? Do they die in patches? Do subtypes die together, or is there a relationship between two or more of the multiple photoreceptor cell types? These questions can be answered by using two labels: one marker for the cell type of interest, and one for apoptotic cells.

Patterns of cell birth (mitosis) also occur within neural tissues. Neurons are not born randomly, and neuronal precursor cells do not divide randomly. Glial and other accessory cells divide during certain diseases or in cases of injury or trauma to the CNS, and markers for dividing cells can be used to image these patterns. Although difficult, it is possible to use multiple specific labels to determine patterns of birth and death for specific subtypes of cells.

The above examples illustrate a clear and urgent need for data management and information discovery tools in the context of biomolecular images. The imaging procedures produce an enormous number of images that are generally analyzed visually, image by image, one by one. A full understanding of the observed behaviors and interrelationships among the proteins is severely limited by the manual mode of data analysis, especially with respect to detecting spatial or temporal patterns of behavior among the many imaged molecules. Rigorous quantitative analyses will surely provide mechanistic insights into normal cell behavior and various neurodegenerative disease processes. This understanding will drive the development of *in silico* models which in turn will direct the *in vivo* experiments. Significant progress toward automating fluorescence image analysis of subcellular patterns has already been made. Sets of subcellular location features that can be used to distinguish all major subcellular structures in both 2D [3, 27] and 3D images [34] have been developed. A critical finding is that automated methods are able to discriminate images of two Golgi proteins that cannot be distinguished by human observers [28]. The work clearly demonstrates the promise of pattern recognition techniques to this domain. The future challenges are to apply these methods to much larger sets of images, improving them as necessary, and to extend them to time series, multispectral and multimodal images.

The specific information processing and data management challenges in the development of the above biomolecular image libraries include:

- Supporting complex queries on new types of data. This implies the development of appropriate data models, query primitives, and index structures.

- Integrating heterogeneous distributed data. The data integration needs to happen over distributed, heterogenenous databases and also over different image modalities such as AFM and fluorescence imaging.

- Supporting interpreted information. A flexible and extensible system that can support biological tools and different layers of interpretation is needed. This is true for biological data in general and for bioimages in particular.

3. MODELS AND QUERIES

Biomolecular images have a high processing and storage cost. A 2D protein localization image from confocal microscopy can require 4 MB (1M pixels recorded in two channels) of storage. A 3D localization image can be 200 MB (50 z-slices). A time series of 50 such 3D localizations that record dynamic information can be 10 GB. This is the result of acquisition from a single sample, and typical experiments involve dozens of samples for different proteins or under different conditions. AFM images require even larger amounts of storage capacity. But it is not just the needed storage that makes the problem of designing bioimage databases daunting. The images have to be analyzed, and visual descriptions extracted using image processing tools (manually or automatically), and these extracted metadata have to be associated with other sources of biological data such as genomics and proteomics [4, 19]. This analysis can lead to a multifold increase in the amount of storage and complexity. Clearly, the amount of information to be maintained and accessed in such a bioimage database is enormous.

Effective description and management of high-throughput experimental data and their relationship to other biological data is critical in the post-genomic era. As compared to traditional scientific databases, typical analyses in biology are much more complex as value is added through the close association of specific data resources. Clear and intuitive models for biological data, particularly for those derived from image data, can be surprisingly challenging. However, a good data model that is sensitive to the novel characteristics, semantics, and diversity will allow the information to be stored, queried, mined, and used effectively. This will not only allow information discovery to happen through a combination of descriptive sources with experimental observations but also the development of mathematical models based on *in vivo* experiments. Such models will in turn permit important questions regarding biological processes to be investigated *in silico* and through more effective *in vivo* experiments.

Queries in a bioimage database can be divided into four classes based on the degree of semantics and interpretation.

- Metadata queries. These are basic queries on the metadata associated with the bioimages.
- Spatial queries. These are queries on the spatial features extracted from the bioimages. For example, images with a subcellular pattern similar to a query image can be found by extracting texture features and using a suitable distance metric.
- Semantic queries. These queries are based on high-level semantic objects, such as cell types, that are extracted from the bioimages manually or automatically.
- Spatio-temporal queries. These queries consider the spatio-temporal changes of features and high-level objects such as protein localization or cytoskeleton growth.

Next, we give more details on each kind of query.

3.1 Metadata queries

Typical metadata fields from the experiments will be *date, scientist, lab, experimental setup, microscope, light sources, filters, camera, experiment, species, antibodies for each channel, and experimental conditions (e.g., normal retina, retina detached for N days, retina reattached for N days, retina under increased oxygen concentration)*. Some specific queries in this class are as follows.

- Find all images from the same experiment as a given image ID.
- Find all experiments that contain normal cat images that have been labeled with calretinin (a calcium-binding protein) both under normal conditions and after 3 days of retinal detachment.

Answering this class of queries is relatively straightforward using current database engines once an appropriate database schema has been developed.

3.2 Spatial queries

Simple spatial features based on texture and shape can be extracted from the images. This can be done at multiple spatial resolutions to provide more flexibility for querying and browsing. The most important task will be to define the right metrics for comparing images based on the extracted features, especially since the images will be produced under different experimental conditions and will be of different subjects. The distance metrics will also need to be supplemented with a statistical model that defines the distribution of the distances. Finally, the extracted features will be high-dimensional and one is faced with the usual challenges of content-based search in such spaces [11, 20, 22].

Some typical queries in this class are as follows:

- Find all images in which vimentin (a filament protein) has a spatial distribution similar to that in a given image.
- Find all pairs of images from the same experiment in which the distribution of vimentin changes as a result of detachment.
- Find all pairs of AFM images that contain a similar texture.
- Find all AFM images that contain patterns similar to a user-specified AFM image illustrating the binding of annexin VI (a calcium-binding protein) to a membrane.

3.3 Semantic queries

Queries in this class are based on semantics extracted from the images. Typical examples of such semantics are the types of cells, their shapes, and their relative location. Semantics can be extracted manually or automatically. This process will be eased through an *atlas* that define the expected distribution of cells under different experimental conditions.

Some examples of queries in this class are as follows:

- Find all normal retinal cell images that contain horizontal cells.
- Find all retinal images of Muller cells labeled by vimentin and GFAP (Glial Fibrillary Acidic Protein).
- Find all images that show Muller cells in which the distribution of CD44 protein is abnormal.

The addition of semantics or interpretations to the content of databases raises a number of issues. How are the interpretations stored and queried? How is the hierarchy of interpretations structured? How is information regarding the accuracy of the interpretations stored and used? How is provenance tracked? These are some of the database design questions that need to be answered.

3.4 Spatio-temporal queries

Spatio-temporal queries consider the time-based evolution of cells and disease processes. Supporting such queries in a meaningful manner requires the extraction of appropriate temporal information from a set of images. The system should provide tools for the modeling of cell behaviors, changes in protein localizations, and disease processes. Queries will typically examine correlations between sets of images or across images and cell/disease models.

The temporal aspect can be observed either by conducting an experiment at different time intervals (e.g., studying retinal images detached for different lengths of time), or by directly observing a change (e.g., movement of a microtubule). In the latter case, temporal features will be useful for an individual microtubule and also for groups of microtubules in order to understand their collective behavior.

Some examples of spatio-temporal queries are as follows:

- Find all image datasets in which the change of vimentin within Muller cells is similar to that observed in the change of GFAP.
- Find patterns of apoptotic cell death within cell populations for a given set of images, and then search for similar patterns.

4. EXTRACTION OF FEATURES

Similarity queries on images has been an active area of research in recent years. Typically, the database images are processed to extract "interesting" descriptors that characterize visual features such as texture and shape. A distance metric is defined that allows similarity comparisons. Nearest neighbors in the feature space using such a metric are expected to match the visual similarity of the corresponding images. Past research on texture features [14, 25] for similarity based search and retrieval, initially developed for aerial images, can be adapted to molecular image databases. Whereas texture features can characterize region properties, statistical shape features can help in the analysis of more structured patterns [8, 10]. Specific challenges facing the extraction of the visual features are in recognizing patterns formed by proteins in varying distribution of proteins within cells with high degree of variability in size, shape and orientation. Previously, sets of *Subcellular Location Features* have been defined that are insensitive to these sources of variation but are still capable of capturing the essence of protein patterns [3, 27, 34]. This has been demonstrated by using them to build image classifiers that can recognize all major subcellular patterns in single cultured cells, both in 2D images [3] and 3D images [34]. Future challenges include applying these approaches to more complex images of multiple cells or tissues in which determination of cell boundaries (segmentation) is required. Automated and semi-automated image segmentation techniques [30, 31] will also be useful in this context.

Low-level visual features are useful in similarity-based retrieval tasks, but cannot answer queries about specific objects or events in images such as the presence of certain proteins. This so-called *semantic gap* is due to the simple fact that the descriptors in the feature space and their nearest neighbors in that space may not correspond well to the perceived visual similarity of the objects. Learning algorithms can help in bridging the semantic gap between low-level features and the associated high-level semantics.

Since in many cases the number of distinct patterns expected may not be known, an important pattern recognition problem that needs to be addressed is that of high-dimensional clustering. Typical image descriptors, such as the *Subcellular Location Features* mentioned above, are high-dimensional feature vectors with dimensionality ranging from a few tens to a few hundreds. Developing efficient clustering methods for such feature vectors is important for information discovery in biomolecular image databases. An example is in grouping proteins by high-resolution location pattern for cataloging purposes [6]. The problem of clustering is well studied in the pattern recognition literature [9]. The research in this area has four major thrusts: (1) raw data clustering; (2) discriminative classification, regression, and detection methods; (3) interactive analysis; and (4) hierarchical techniques for the analysis of the large, high-dimensional datasets that arise from high-dimensional visual features.

The high-dimensional aspect of visual features is challenging because conventional data clustering techniques do not scale well with data size and dimensionality. It also poses challenges for indexing: As the dimensionality of the data increases, the cost of searching the database becomes linear even with sophisticated tree-

structured indexes. In the context of database search, new techniques have been developed to support search in high-dimensional datasets specifically when relevance feedback is used [17, 33]. Furthermore, the underlying feature extraction methods can be statistically modeled to provide dimensionality reduction, such as modifying the conventional Gabor texture descriptor to nearly half the size and retaining comparable retrieval performance [2]. Alternative dimensionality reduction techniques include linear transformations as in FastMap [12] and non-linear transformations such as non-linear axis scaling [35]. Application of a number of dimensionality reduction methods to the Subcellular Location Features led to the identification of a set of only eight features (out of over 80) that can be used to recognize all major subcellular patterns with over 86 % accuracy on single cells [16]. While better performance can be achieved with more features, the smaller set is suitable for database search and indexing.

5. CONCLUDING REMARKS

Two significant problems encountered in initial efforts to create biological image databases were the absence of standards for describing samples and image acquisition settings and the diversity of image formats used by microscope manufacturers. The earliest published work on the subject was the initial description of the BioImage database [5], but little progress was made on the goals of the project until recently. Other efforts include the OME project [1] and the PSLID database [15]. There has been some convergence on the desired characteristics of a microscope image database schema, and extensive work on image import has been done for OME. The most recent version of OME [32] addresses a number of problems with the initial release, and is an excellent base for microscopy image informatics efforts.

The focus of this article has been on describing the growing importance of terabyte-scale image collections in cell and molecular biology research and on identifying information technology and machine learning challenges that must be addressed in order to maximize knowledge creation from these collections. We have described relevant preliminary work that shows the feasibility of creating tools to address these challenges in order to advance microscopy from a subjective, descriptive practice based on visual interpretation to an objective, systematic science that can provide critical knowledge on the spatial and temporal patterns of biological macromolecules. It is anticipated that these tools will provide a critical capability for systems biology efforts whose goal is to understand the mechanisms by which all biologically important molecules interact to accomplish their roles at the cell, tissue and organism level.

6. ACKNOWLEDGMENTS

The research work described above is supported by the National Science Foundation under an ITR project titled *Next Generation Bio-Molecular Imaging and Information Discovery* (grant number EF-0331697). Other investigators include Sanjoy Banerjee, Stuart Feinstein, Steven Fisher, S. Jammalamadaka, Kenneth Rose, Jianwen Su, Yuan-Fang Wang, Leslie Wilson (UCSB), Christos Faloutsos, Jelena Kovacevic, Tom Mitchell (CMU), Arun Majumdar (Berkeley), and Peter Sorger (MIT). More details on the project can be found at `http://www.bioimage.ucsb.edu`.

7. REFERENCES

[1] P. D. Andrews, I. S. Harper, and J. R. Swedlow. To 5D and beyond: Quantitative fluorescence microscopy in the postgenomic era. *Traffic*, 3(1):29–36, Jan. 2002.

[2] S. Bhagavathy, J. Tesic, and B. Manjunath. On the Raleigh nature of Gabor filter outputs. In *Proc. ICPR*, 2003.

[3] M. V. Boland and R. F. Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics*, 17(12):1213–1223, 2001.

[4] O. Camoglu, T. Kahveci, and A. K. Singh. Towards index-based similarity search for protein structure databases. In *Proc. IEEE Computer Society Bioinformatics Conference*, 2003.

[5] J. Carazo and E. H. Stelzer. The BioImage database project: Organizing multidimensional biological images in an object-relational database. *J. Struct. Biol.*, 125(2–3):97–102, 1999.

[6] X. Chen, M. Velliste, S. Weinstein, J. W. Jarvik, and R. F. Murphy. Location proteomics—building subcellular location trees from high resolution 3D fluorescence microscope images of randomly-tagged proteins. In *Proc. SPIE*, volume 4962, 2003.

[7] N. Cuenca, P. Dong, K. Linberg, S. Fisher, and H. Kolb. Choline acetyltransferase is experessed by non-starburst amacrine cells in the ground squirrel retina. *Brain Res.*, 964:21–30, 2003.

[8] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, 1998.

[9] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.

[10] N. Duta and M. Sonka. Segmentation and interpretation of MR brain images: An improved active shape model. *IEEE Transactions on Medical Imaging*, 17(6):1049–1062, Dec. 1998.

[11] C. Faloutsos. *Searching Multimedia Databases by Content*. Kluwer, 1996.

[12] C. Faloutsos and K.-I. Lin. FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proc. SIGMOD*, pages 163–174, 1995.

[13] R. N. Fariss, R. S. Molday, S. K. Fisher, and B. Matsumoto. Evidence from normal and degenerating photoreceptors that two outer segment integral membrane proteins have separate transport pathways. *J. Comp. Neurol.*, 387:148–156, 1997.

[14] G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE*

Transactions on Image Processing, 8(2):255–269, Feb. 1999.

[15] K. Huang, J. Lin, J. A. Gajnak, and R. F. Murphy. Image content-based retrieval and automated interpretation of flourescence microscope images via the protein subcellular location image database. In *Proc. ISBI*, pages 325–328, July 2002.

[16] K. Huang, M. Velliste, and R. F. Murphy. Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. In *Proc. SPIE*, pages 307–318, 2003.

[17] Y. Ishikawa, R. Subramanya, and C. Faloutsos. MindReader: Querying databases through multiple examples. In *Proc. VLDB*, pages 218–227, 1998.

[18] J. Jarvik, G. Fisher, C. Shi, L. Hennen, C. Hauser, S. Adler, and P. Berget. In vivo functional proteomics: Mammalian genome annotation using CD-tagging. *BioTechniques*, 33:852–867, 2002.

[19] T. Kahveci and A. Singh. MAP: Searching large genome databases. In *Proc. Pacific Symposium on Biocomputing*, 2003.

[20] K. V. R. Kanth, D. Agrawal, and A. K. Singh. Dimensionality reduction for similarity searching in dynamic databases. In *Proc. SIGMOD*, pages 166–176, 1998.

[21] A. Kumar, S. Agarwal, J. Heyman, S. Matson, M. Heidtman, S. Piccirillo, L. Umansky, A. Drawid, R. Jansen, Y. Liu, K.-H. Cheung, P. Miller, M. Gerstein, G. Roeder, and M. Snyder. Subcellular localization of the yeast proteome. *Genes Dev.*, 16:707–719, 2002.

[22] C. A. Lang and A. K. Singh. Modeling high-dimensional index structures using sampling. In *Proc. SIGMOD*, pages 389–400, 2001.

[23] G. Lewis, K. Linberg, and S. Fisher. Neurite outgrowth from bipolar and horizontal cells after experimental retinal detachment. *Invest. Opthal. Vis. Sci.*, 39(2):424–34, Feb. 1998.

[24] G. Lewis, B. Matsumoto, and S. Fisher. Changes in the organization and expression of cytoskeletal proteins during retinal degeneration induced by retinal detachment. *Invest. Opthal. Vis. Sci.*, 36:2404–2416, 1995.

[25] B. S. Manjunath, P. Salembier, and T. Sikora, editors. *Introduction to MPEG 7: Multimedia Content Description Language*. Wiley, 2002.

[26] R. Marc, R. Murry, S. Fisher, K. Linberg, and G. Lewis. Amino acid signatures in the detached cat retina. *Invest. Opthal. Vis. Sci.*, 39:1694–1702, 1998.

[27] R. F. Murphy, M. Velliste, and G. Porreca. Robust classification of subcellular location patterns in fluorescence microsocope images. In *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 67–76, 2002.

[28] R. F. Murphy, M. Velliste, and G. Porreca. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *J. VLSI Sig. Proc.*, 35:311–321, 2003.

[29] M. Rolls, P. Stein, S. Taylor, E. Ha, F. McKeon, and T. Rapoport. A visual screen of a GFP-fusion library identifies a new type of nuclear envelope membrane protein. *J. Cell Biol.*, 146:29–44, 1999.

[30] M. E. Saban and B. Manjunath. Video region segmentation by spatio-temporal watersheds. In *Proc. ICIP*, Barcelona, Spain, Sept. 2003.

[31] B. Sumengen, B. Manjunath, and C. Kenney. Image segmentation using multi-region stability and edge stretch. In *Proc. ICIP*, pages 429–432, Sept. 2003.

[32] J. R. Swedlow, I. Goldberg, E. Brauner, and P. K. Sorger. Informatics and quantitative analysis in biological imaging. *Science*, 300:100–102, 2003.

[33] J. Tešić and B. Manjunath. Nearest neighbor search for relevance feedback. In *Proc. CVPR*, volume 2, June 2003.

[34] M. Velliste and R. F. Murphy. Automated determination of protein subcellular locations from 3D fluorescence microscope images. In *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging*, pages 867–870, July 2002.

[35] L. Wu and C. Faloutsos. Making every bit count: fast nonlinear axis scaling. In *Proc. KDD*, pages 664–669. ACM Press, 2002.