

Evaluating Classifiers

- Divide ~100 images for each class into **training** set and **test** set
- Use the **training** set to determine rules for the classes
- Use the **test** set to evaluate performance
- Repeat with different division into training and test
- Evaluate different sets of features chosen as most discriminative by feature selection methods
- Evaluate different classifiers (NN, SVM, MOE)

Carnegie Mellon

2D HeLa Classification Results

True Class	Output of the Classifier									
	DNA	ER	Gia	Gpp	Lam	Mit	Nuc	Act	TfR	Tub
DNA	99	1	0	0	0	0	0	0	0	0
ER	0	97	0	0	0	2	0	0	0	1
Gia	0	0	91	7	0	0	0	0	2	0
Gpp	0	0	14	82	0	0	2	0	1	0
Lam	0	0	1	0	88	1	0	0	10	0
Mit	0	3	0	0	0	92	0	0	3	3
Nuc	0	0	0	0	0	0	99	0	1	0
Act	0	0	0	0	0	0	0	100	0	0
TfR	0	1	0	0	12	2	0	1	81	2
Tub	1	2	0	0	0	1	0	0	1	95

Overall accuracy = 92%

Carnegie Mellon

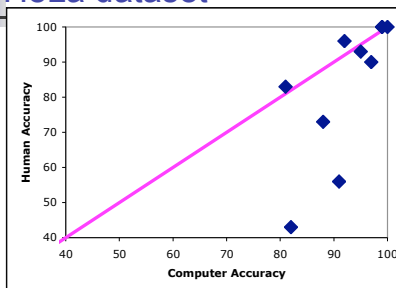
Human Classification Results

True Class	Output of the Classifier									
	DNA	ER	Gia	Gpp	Lam	Mit	Nuc	Act	TfR	Tub
DNA	100	0	0	0	0	0	0	0	0	0
ER	0	90	0	0	3	6	0	0	0	0
Gia	0	0	56	36	3	3	0	0	0	0
Gpp	0	0	54	33	0	0	0	0	3	0
Lam	0	0	6	0	73	0	0	0	20	0
Mit	0	3	0	0	0	96	0	0	0	3
Nuc	0	0	0	0	0	0	100	0	0	0
Act	0	0	0	0	0	0	0	100	0	0
TfR	0	13	0	0	3	0	0	0	83	0
Tub	0	3	0	0	0	0	0	3	0	93

Overall accuracy = 83% (92% for major patterns)

Carnegie Mellon

Computer vs. Human on 2D HeLa dataset



Carnegie Mellon

3D Morphological Features (SLF-9)

- 28 features, 14 from protein objects and 14 from their relationship to corresponding DNA images
 - Based on number of objects, object size, object distance to COF
- Corresponding DNA image required
- A subset of 9 features selected by SDA forms SLF10

Carnegie Mellon

SLF-14

- 14 SLF-9 features that do not require DNA images
- 2 Edge features
 - Ratio of above threshold pixel along an edge
 - Ratio of fluorescence along an edge
- 26 3D Haralick texture features
 - GLCM built on 13 directions
 - One set (13) of mean features and the other set (13) of range features

Carnegie Mellon

Pixel Resolution and Gray Levels

- Texture features are potentially influenced by the number of gray levels and pixel resolution of the image
- Optimization for each image dataset required

Carnegie Mellon

SLF-17

- A feature subset with 7 features selected from SLF-14 at 256 gray levels and 0.4 micron pixel resolution
 - 1 morphological feature
 - 1 edge feature
 - 5 texture features

Carnegie Mellon

3D HeLa Classification Results

True Classes	Output of the Classifier									
	DNA	ER	Gia	Gpp	Lam	Mit	Nuc	Act	TfR	Tub
DNA	98	2	0	0	0	0	0	0	0	0
ER	0	100	0	0	0	0	0	0	0	0
Gia	0	0	100	0	0	0	0	0	0	0
Gpp	0	0	0	96	4	0	0	0	0	0
Lam	0	0	0	4	95	0	0	0	0	2
Mit	0	0	2	0	0	96	0	2	0	0
Nuc	0	0	0	0	0	0	100	0	0	0
Act	0	0	0	0	0	0	0	100	0	0
TfR	0	0	0	0	2	0	0	0	96	2
Tub	0	2	0	0	0	0	0	0	0	98

Overall accuracy = 98%

Carnegie Mellon

Next generation goals

- Have demonstrated
 - discrimination of patterns that cannot be distinguished by visual examination
 - discrimination is better with 3D images than 2D images
- New goal is to assign proteins to **high-resolution** location categories that are learned rather than predefined
- New goal: use these methods to classify images in **web sources** (e.g., **journal articles**)

Carnegie Mellon

Clustering by Image Similarity

- SLF features can be used to measure similarity of protein patterns
- This allows us for the first time to create a systematic, objective, framework for describing subcellular locations
- Ideal for database references
- One way is by creating a **Subcellular Location Tree**

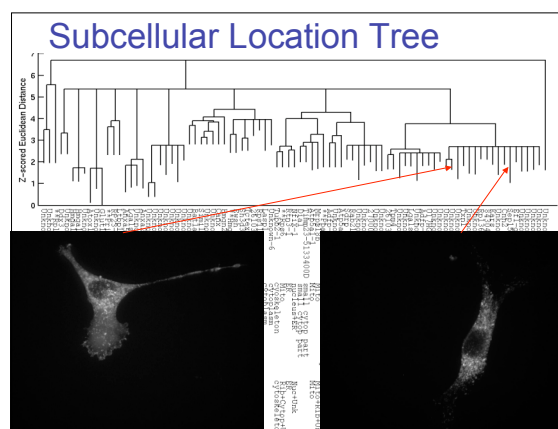
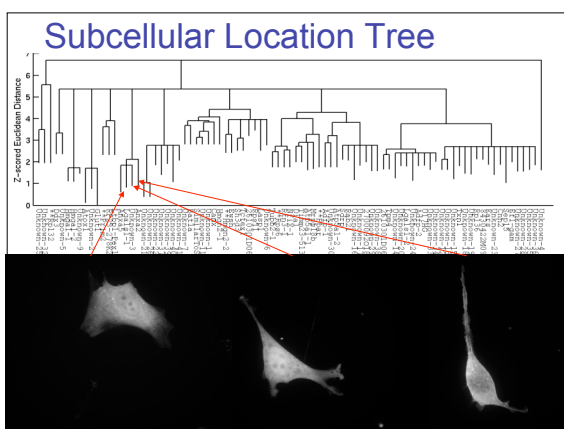
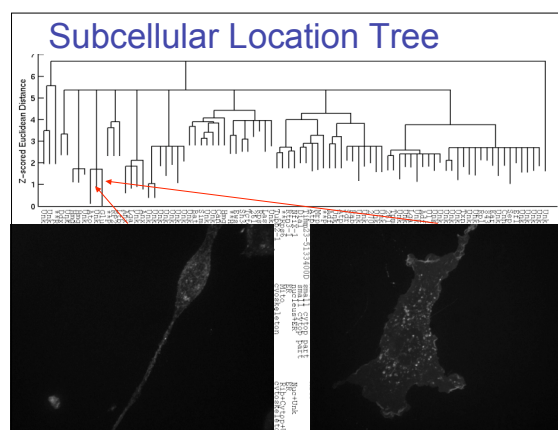
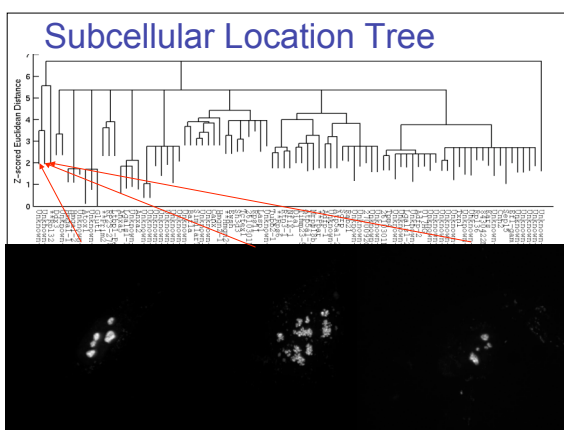
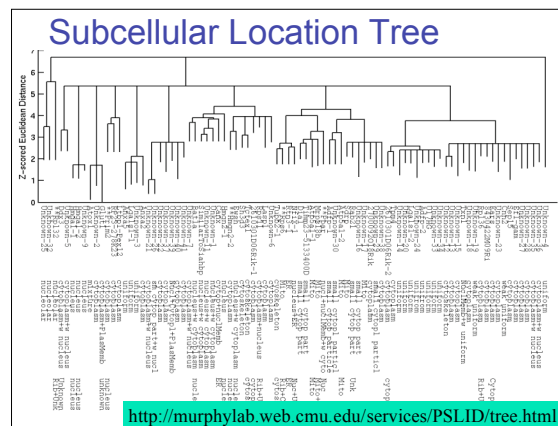
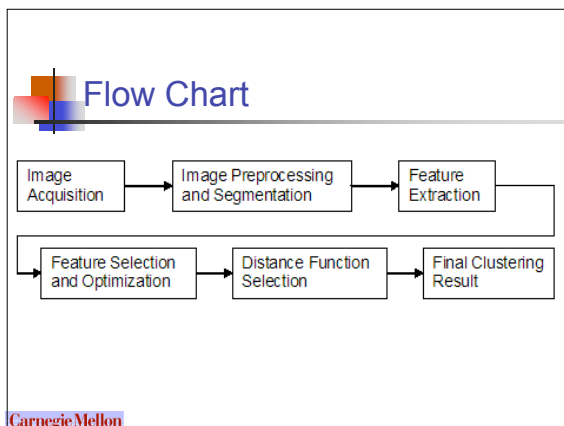
Carnegie Mellon

Clustering by Image Similarity

- Analyze 3D 3T3 dataset (90 clones obtained by CD-tagging)
- Use resampling to estimate high confidence tree structure
- Draw as best-fit pseudo-additive tree

Carnegie Mellon

(c) Murphy and Faloutsos, 2005



Fundamental issue

- How do we know whether a particular Subcellular Location Tree is the correct one?
- Things to vary that change tree
 - Clustering method
 - Distance function
- Evaluation - objective function

Carnegie Mellon

Significance

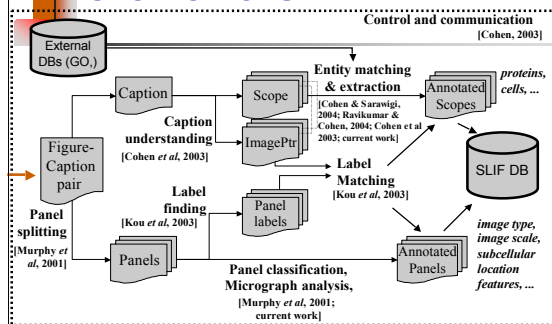
- Can **subdivide** clusters by observing response to drugs, oncogenes, etc.
- These represent protein location **states**
- Base knowledge required for **modeling** (systems biology)
- Can be used to identify potential protein **interactions**

Carnegie Mellon

C. Image Content-based Retrieval and Interpretation of Micrographs from On-line Journal Articles

Carnegie Mellon

Overview of SLIF



Carnegie Mellon

Overview: Image processing in SLIF

(1)
Segment
into
“panels”

Detect & remove
annotations

Classify
panels

A **U2B"-GFP** **anti-collin AB**

B **U2B"-GFP** **anti-U2B" AB**

FMI+ **FMI+**

Carnegie Mellon

Overview: Image processing in SLIF

(2) Find **scalebar** and **scale measurement**

Rescale image of each cell, adjust **contrast**, and compute subcellular localization **features** as if it were an ordinary microscope image.

Of course, you still don't know what it's an image *of*...



Carnegie Mellon

(c) Murphy and Faloutsos, 2005

Overview: Text Processing in SLIF

- Find entity names in text, and panel labels in text and the image.
- Match panels labels in text to panel labels on the image.
- Associate entity names to textual panel labels using *scoping* rules.

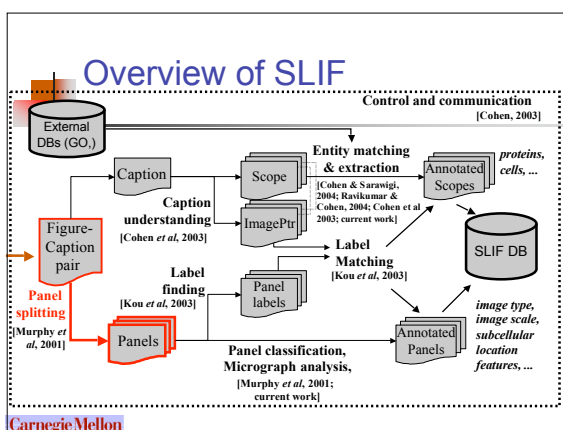
Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B'-GFP, double labeled with GFP (left panel) and anti-coilin AB (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B' 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B' 0 also express the U2B' 0-GFP fusion. Bars, 5 m m. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B'-GFP, double labeled with GFP (left panel) and anti-coilin AB (right panel). Three nuclei are shown....

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B'-GFP, double labeled with GFP (left panel) and anti-coilin AB (right panel). Three nuclei are shown....

Carnegie Mellon



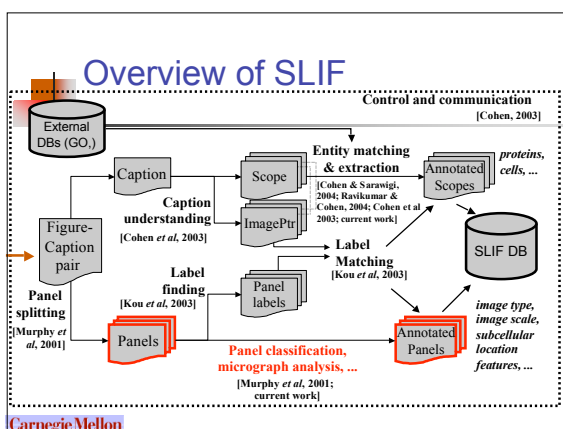
Panel-Splitting

[Murphy et al, BIBE 2001]

- 1) Compute avg horizontal & vertical intensity
- 2) Split at high contrast narrow bands
- 3) Recurse if necessary

New methods improve on 74% precision, 60% recall of 2001 article

Carnegie Mellon



Panel-level processing in SLIF

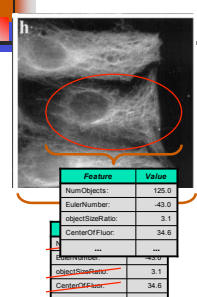
[Murphy et al, Proc. BIBE 2001]

- Classification:** is this a FMI, or something else?
 - 97% precision, 92% recall using gray-level histogram and K-NN
- Annotation removal:** discard artificial markup
 - 83% precision, 82% recall
- Scale-finding:** find scale bar and stated size in μm
 - 76% precision, 50% recall, improvements being explored
- Segmentation:** where are the cell boundaries?
 - hard! now finessed

Carnegie Mellon

Panel-level processing in SLIF

[current work, Murphy & Juchang Hua]



Feature	Value
NumObjects	125.0
EulerNumber	-43.0
AspectRatioRatio	3.1
CenterOffFactor	34.6
...	...
ObjectSizeFactor	3.1
CenterOffFactor	34.6
...	...

- Segmentation:** where are the cell boundaries?
 - Murphy *et al.*, 2001: explicit segmentation is difficult
 - New approach: encode localization with features computed on a *field* containing several different *cells*
 - Are these features equally predictive?

Carnegie Mellon

Cell-level vs field-level features

[current work, Murphy & Juchang Hua]

True	Cell-level features										Output of the Classifier
Classification	DNA	ER	Giant	gpp	LAMP	Mito.	Nucle.	Actin	TfR	Tubul.	
DNA	99%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
ER	0%	89%	0%	0%	4%	4%	0%	0%	1%	2%	
Giantin	0%	0%	76%	20%	0%	1%	1%	0%	1%	0%	
gpp130	0%	0%	23%	73%	0%	1%	2%	0%	1%	0%	
LAMP2	0%	2%	0%	0%	83%	1%	0%	0%	13%	0%	
Mitoch.	0%	5%	0%	0%	2%	90%	0%	0%	1%	2%	
Nucleolin	0%	0%	0%	0%	0%	0%	98%	0%	0%	0%	
Actin	0%	0%	0%	0%	0%	0%	0%	99%	0%	1%	
TfR	0%	3%	0%	0%	16%	3%	0%	1%	75%	2%	
Tubulin	0%	2%	0%	0%	0%	2%	0%	0%	3%	93%	

Carnegie Mellon

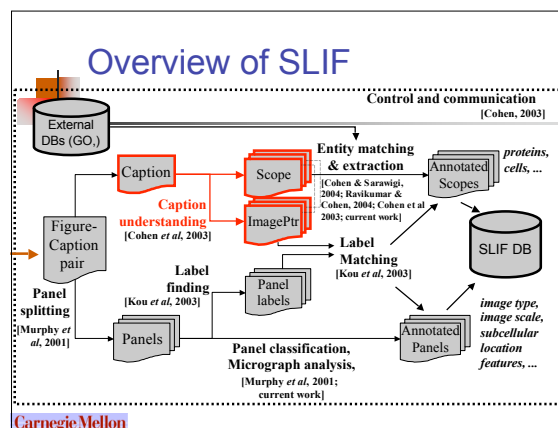
Cell-level vs field-level features

[Huang & Murphy (2004)]

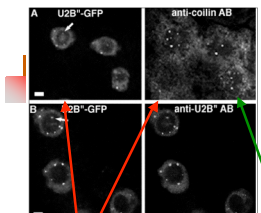
Field-level features—show comparable performance

	DNA	ER	Giant.	Gpp	LAMP2	Mito.	Nucl.	Actin	TfR	Tubul.
DNA	100	0	0	0	0	0	0	0	0	0
ER	0	100	0	0	0	0	0	0	0	0
Giantin	2	0	68	26	4	0	0	0	0	0
Gpp130	0	0	44	56	0	0	0	0	0	0
LAMP2	0	0	0	0	82	2	0	0	16	0
Mitochondria	2	0	0	0	0	96	0	0	0	2
Nucleolin	4	0	0	0	0	0	96	0	0	0
Actin	0	0	0	0	0	2	0	98	0	0
TfR	4	0	0	0	18	4	0	2	68	4
Tubulin	2	8	0	0	0	8	0	0	10	72

Carnegie Mellon



Text Processing in SLIF

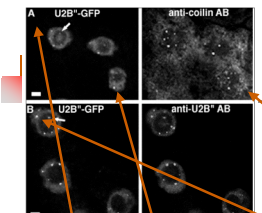


- Find entity names in text, and **panel labels** in text and the image.
- Match panels labels in text to panel labels on the image.
- Associate entity names to textual panel labels using **scoping** rules.

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and autoantibody against p80 coilin (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 μm. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

Text Processing in SLIF



Identify all **image pointers**: Substrings that refer to parts of the image

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and autoantibody against p80 coilin (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 μm. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

(c) Murphy and Faloutsos, 2005

Classify image pointers as citation-style or bullet-style.

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and autoantibody against p80 coilin (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 m m. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

Style determines scope:

- The scope of a **bullet-style** image pointer is all words between it and the next "bullet"

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and autoantibody against p80 coilin (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 m m. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

Style determines scope:

- The scope of a **bullet-style** image pointer is all words between it and the next "bullet"
- The scope of a **citation-style** image pointer is some set of words nearby it (heuristically determined by separating words and punctuation)

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and autoantibody against p80 coilin (right panel). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anti-coilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0-GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 m m. Movement of Coiled Bodies Vol. 10, July 1999 2299

Carnegie Mellon

SLIF Service - Microsoft Internet Explorer

Address: http://greenlab.mbc.cmu.edu:8080/home.jsp

Murphy lab SLIF Service

Home
Search by caption
Search by resolution
Search by protein name
Search FMI
Search results

SLIF stands for Subcellular Location Image Finder, which automatically extracts information about protein subcellular locations from biological literature. By looking at the figure-caption pairs in the paper, SLIF applies image processing methods to analyze fluorescence microscope images and extract a quantitative description of the localization pattern of the tagged proteins. Detailed textual annotations of the images and associated proteins are also generated by analyzing the accompanying captions.

Our long-term goal is to develop a large library of annotated and analyzed fluorescence microscope images, in order to support data mining.

Carnegie Mellon

SLIF Service - Microsoft Internet Explorer

Address: http://greenlab.mbc.cmu.edu:8080/search_caption.jsp

Murphy lab SLIF Service

Home
Search by caption
Search by resolution
Search by protein name
Search FMI
Search results

Search for:

Maximum number of displayed records: 100

Search

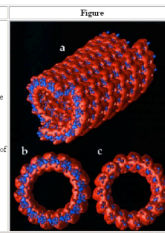
Carnegie Mellon

SLIF Service - Microsoft Internet Explorer

Address: http://greenlab.mbc.cmu.edu:8080/query_caption.jsp

Murphy lab SLIF Service

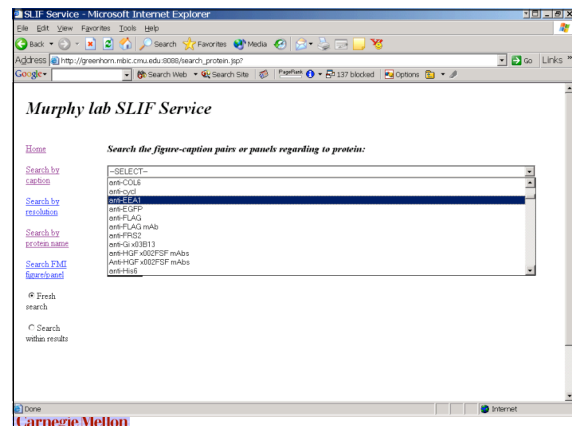
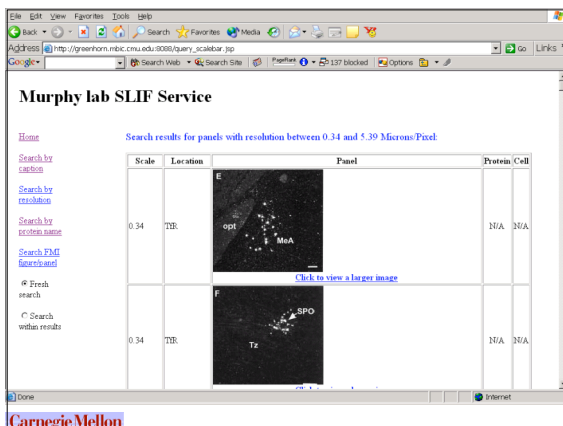
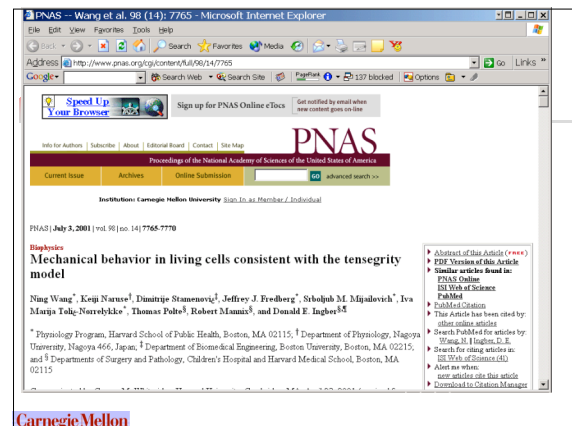
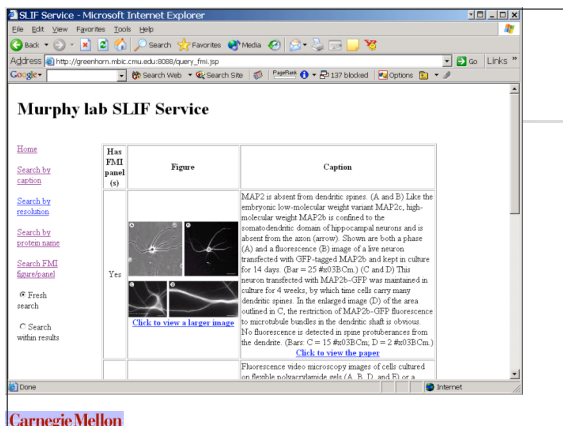
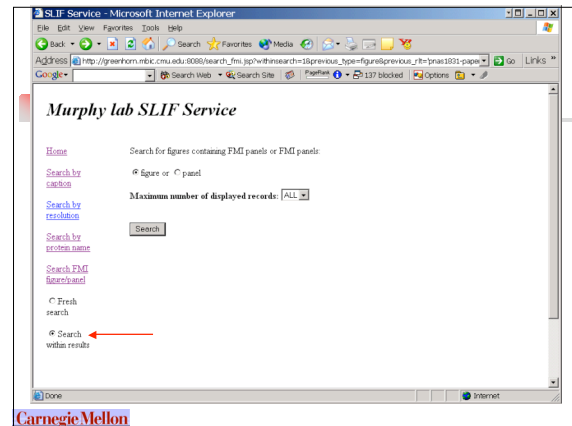
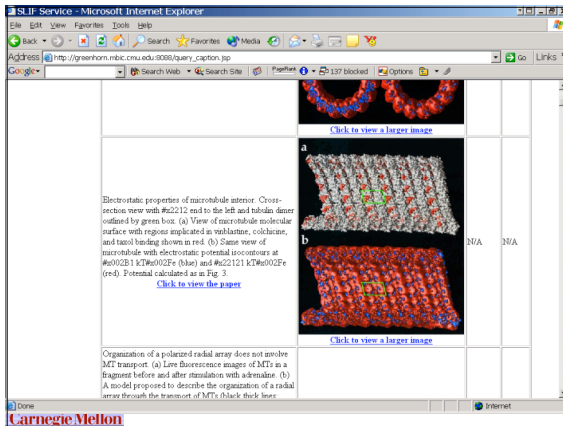
SLIF found 46 caption and displayed 46 matched caption for query "microtubule, mt, tubulin".

Caption	Figure	Protein	Cell
Electron micrographs of the microtubule exterior. Potential protofilaments are shown at #602B1, #602B2, #602B3, #602B4, #602B5, #602B6, #602B7, #602B8, #602B9, #602B10, #602B11, #602B12, #602B13, #602B14, #602B15, #602B16, #602B17, #602B18, #602B19, #602B20, #602B21, #602B22, #602B23, #602B24, #602B25, #602B26, #602B27, #602B28, #602B29, #602B30, #602B31, #602B32, #602B33, #602B34, #602B35, #602B36, #602B37, #602B38, #602B39, #602B40, #602B41, #602B42, #602B43, #602B44, #602B45, #602B46, #602B47, #602B48, #602B49, #602B50, #602B51, #602B52, #602B53, #602B54, #602B55, #602B56, #602B57, #602B58, #602B59, #602B60, #602B61, #602B62, #602B63, #602B64, #602B65, #602B66, #602B67, #602B68, #602B69, #602B70, #602B71, #602B72, #602B73, #602B74, #602B75, #602B76, #602B77, #602B78, #602B79, #602B80, #602B81, #602B82, #602B83, #602B84, #602B85, #602B86, #602B87, #602B88, #602B89, #602B90, #602B91, #602B92, #602B93, #602B94, #602B95, #602B96, #602B97, #602B98, #602B99, #602B100.		#602B1, #602B2, #602B3, #602B4, #602B5, #602B6, #602B7, #602B8, #602B9, #602B10, #602B11, #602B12, #602B13, #602B14, #602B15, #602B16, #602B17, #602B18, #602B19, #602B20, #602B21, #602B22, #602B23, #602B24, #602B25, #602B26, #602B27, #602B28, #602B29, #602B30, #602B31, #602B32, #602B33, #602B34, #602B35, #602B36, #602B37, #602B38, #602B39, #602B40, #602B41, #602B42, #602B43, #602B44, #602B45, #602B46, #602B47, #602B48, #602B49, #602B50, #602B51, #602B52, #602B53, #602B54, #602B55, #602B56, #602B57, #602B58, #602B59, #602B60, #602B61, #602B62, #602B63, #602B64, #602B65, #602B66, #602B67, #602B68, #602B69, #602B70, #602B71, #602B72, #602B73, #602B74, #602B75, #602B76, #602B77, #602B78, #602B79, #602B80, #602B81, #602B82, #602B83, #602B84, #602B85, #602B86, #602B87, #602B88, #602B89, #602B90, #602B91, #602B92, #602B93, #602B94, #602B95, #602B96, #602B97, #602B98, #602B99, #602B100.	3BA

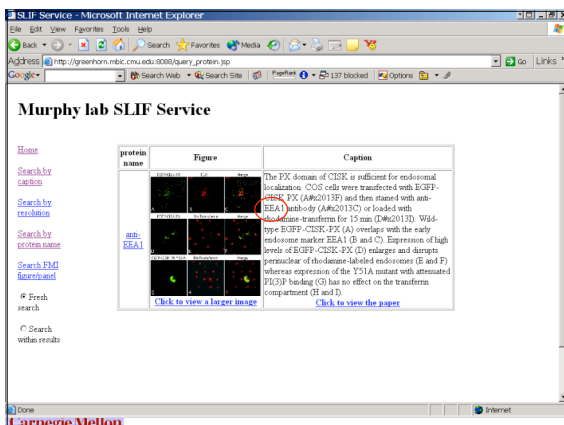
Click to view the pages

Carnegie Mellon




(c) Murphy and Faloutsos, 2005






(c) Murphy and Faloutsos, 2005



Acknowledgments

- Students
 - Dr. Michael Boland
 - Dr. Mia Markey
 - Gregory Porreca (summer student)
 - Dr. Meel Velliste
 - Dr. Kai Huang
 - Xiang Chen
 - Juchang Hua
 - Zhenzhen Kou
- Funding
 - NSF, NIH, Commonwealth of Pennsylvania
- Collaborators/Consultants
 - Simon Watkins, Tom Mitchell, Christos Faloutsos, Jon Jarvik, Peter Berget, Jelena Kovacevic, William Cohen, David Casasent
 - NSF large ITR grant: B. S. Manjunath Ambuj Singh, Peter Sorger

Carnegie Mellon

<http://murphylab.web.cmu.edu/publications>

Review Articles

- Y. Hu and R. F. Murphy (2004). Automated Interpretation of Subcellular Patterns from Immunofluorescence Microscopy. *J. Immunol. Methods* 290:93-105.
- K. Huang and R. F. Murphy (2004). From Quantitative Microscopy to Automated Image Understanding. *J. Biomed. Optics* 9:893-912.
- K. Huang and R. F. Murphy (2004). Data Mining Methods for a Systematics of Protein Subcellular Location. In: **Data Mining in Bioinformatics** (J.T.L. Wang, M.J. Zaki, H.T.T. Toivonen, and D. Shasha, eds.), Springer-Verlag (London), pp. 143-187.

Carnegie Mellon

First published system for recognizing subcellular location patterns - 2D CHO (5 patterns)

- M. V. Boland, M. K. Markey and R. F. Murphy (1997). Automated Classification of Cellular Protein Localization Patterns Obtained via Fluorescence Microscopy. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 594-597.
- M. V. Boland, M. K. Markey and R. F. Murphy (1998). Automated Recognition of Patterns Characteristic of Subcellular Structures in Fluorescence Microscopy Images. *Cytometry* 33:366-375.

Carnegie Mellon

2D HeLa pattern classification (10 major patterns)

- R. F. Murphy, M. V. Boland and M. Velliste (2000). Towards a Systematics for Protein Subcellular Location: Quantitative Description of Protein Localization Patterns and Automated Analysis of Fluorescence Microscope Images. *Proc Int Conf Intell Syst Mol Biol* 8:251-259.
- M. V. Boland and R. F. Murphy (2001). A Neural Network Classifier Capable of Recognizing the Patterns of all Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells. *Bioinformatics* 17:1213-1223.

Carnegie Mellon

3D HeLa pattern classification (11 major patterns)

- M. Velliste and R.F. Murphy (2002). Automated Determination of Protein Subcellular Locations from 3D Fluorescence Microscope Images. *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI 2002)*, pp. 867-870.

Carnegie Mellon

Improving features, feature selection, classification method

- R.F. Murphy, M. Velliste, and G. Porreca (2003). Robust Numerical Features for Description and Classification of Subcellular Location Patterns in Fluorescence Microscope Images. *J. VLSI Sig. Proc.* 35: 311-321.
- K. Huang, M. Velliste, and R. F. Murphy (2003). Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. *Proc. SPIE* 4962: 307-318.

Carnegie Mellon

Improving features, feature selection, classification method

- K. Huang and R.F. Murphy (2004). Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *BMC Bioinformatics* 5:78.
- X. Chen and R.F. Murphy (2004). Robust Classification of Subcellular Location Patterns in High Resolution 3D Fluorescence Microscope Images. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1632-1635.

Carnegie Mellon

Subcellular Location Trees - 3D 3T3 CD-tagged images

- X. Chen, M. Velliste, S. Weinstein, J.W. Jarvik and R.F. Murphy (2003). Location proteomics - Building subcellular location trees from high resolution 3D fluorescence microscope images of randomly-tagged proteins. *Proc. SPIE* 4962: 298-306.
- X. Chen and R. F. Murphy (2005). Objective Clustering of Proteins Based on Subcellular Location Patterns. *Journal of Biomedicine and Biotechnology*, in press.

Carnegie Mellon

Mining images from journal articles - SLIF

- R. F. Murphy, M. Velliste, J. Yao, and G. Porreca (2001). Searching Online Journals for Fluorescence Microscope Images Depicting Protein Subcellular Location Patterns. *Proceedings of the 2nd IEEE International Symposium on Bio-Informatics and Biomedical Engineering (BIBE 2001)*, pp. 119-128.
- W.W. Cohen, R. Wang and R.F. Murphy (2003). Understanding Captions in Biomedical Publications. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*, pp. 499-504.
- Z. Kou, W.W. Cohen and R.F. Murphy (2003). Extracting information from text and images for location proteomics. *Proceedings of the Third ACM SIGKDD Workshop on Data Mining in Bioinformatics (BIODDD03)*, pp. 2-9.

Carnegie Mellon

Mining images from journal articles - SLIF

- R. F. Murphy, Z. Kou, J. Hua, M. Joffe, W.W. Cohen (2004) Extracting and Structuring Subcellular Location Information from On-Line Journal Articles: The Subcellular Location Image Finder. *Proceedings of the IASTED International Conference on Knowledge Sharing and Collaborative Engineering (KSCE 2004)*, pp. 109-114.
- Z. Kou, W. W. Cohen, and R.F. Murphy (2005) High-recall protein entity recognition using a dictionary. *ISMB 2005*, in press.

Carnegie Mellon

PSLID - Protein Subcellular Location Image Database

- K. Huang, J. Lin, J.A. Gajnak, and R.F. Murphy (2002). Image Content-based Retrieval and Automated Interpretation of Fluorescence Microscope Images via the Protein Subcellular Location Image Database. *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI 2002)*, pp. 325-328.

Carnegie Mellon

(c) Murphy and Faloutsos, 2005



Thank you!



Robert Murphy
murphylab.web.cmu.edu



Christos Faloutsos
www.cs.cmu.edu/~christos

