Classification of Micro Array Protein Crystal Images With Laplacian Pyramidal Filters and Neural Networks

Elsa D. Angelini, Ya Wang, Andrew F. Laine
Heffner Biomedical Imaging Laboratory
Department of Biomedical Engineering, Columbia University, New York, 10027, NY, USA

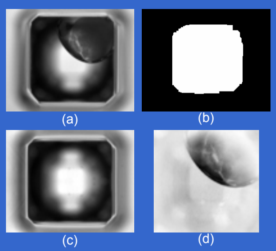
INTRODUCTION

- This project is part of the Northeast Structural Genomics Consortium (NESG). The goal of this consortium is to develop efficient and integrated technologies for high-throughput (HTP) protein production and 3D structure determination.
- This project focuses on the design of an image analysis system to classify protein crystal structures in a production oriented environment.
- The method presented performs classification of microscopic images as protein precipitates or crystals versus clear solutions.
- Preliminary experiments are presented, which provided high classification accuracy on large datasets of protein crystallization experiments using expert classification for ground truth.

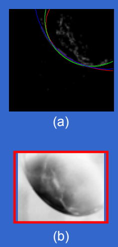
1

METHODOLOGY

1. Preprocessing of Microscopic Images



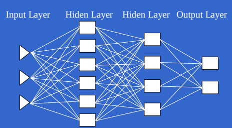
(a) Original image with an oil drop containing a precipitate. (b) Mask computed from the contours of the wells. (c) Background image acquired without drops. (d) Pre-filtered image.



(a) Ellispoidal Hough transform [1] to detect the three most probable ellipses, plotted over the edge map of the pre-filtered image. (b) Pre-filtered image cropped with the minimal rectangular area encompassing the three ellipses.

2. Feature Extraction with Laplacian Pyramidal Expansion [2]

3. Classification with a Feed-Forward Neural Network



Features of the Neural Network: Normalized 10-bins histograms of the Laplacian expansion coefficients from the four first detail levels.

2

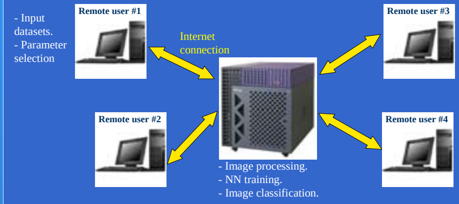
REFERENCES

[1] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, pp. 111-122, 1981.
[2] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, pp. 532-540, 1983.
[3] C. A. Cumbaa, A. Lauricella, N. Fehrman, C. Veatch, R. Collins, J. Luft, G. DeTitta, and I. Jurisica, "Automatic classification of sub-microlitre protein-crystallization trials in 1536-well plates," *Acta Crystallographica Section D-Biological Crystallography*, vol. 59, pp. 1619-1627, 2003.
[4] I. Jurisica, P. Rogers, J. I. Glasgow, R. J. Collins, J. R. Wolfley, J. R. Luft, and G. T. DeTitta, "Improving objectivity and scalability in protein crystallization," *IEEE Intelligent Systems*, vol. 16, pp. 26-34, 2001.

10

METHODOLOGY

Remote Application Server: Web page interface managed by a Matlab© Server.



3

DATABASES

1. Microscopic images were acquired with a CCD camera under robotic control.
2. Gray scale 8-bits images saved in tiff/jpeg format.
3. Arrays of 256 $\times$256 wells screened at three time instants after experimental setup: 1 day , 1 week and 2 weeks.
4. Database of 13,490 scored images:
   - 509 drops containing precipitates and/or crystals.
   - 3,349 clear drops.
   - 9,632 'others' combining drops with organic matter, or dye.

4

INITIAL STUDY

- **Training data set**
  - 100 images
  - 1/3 clear drops, 2/3 drops with precipitates/crystals
- **Binary classification**
  - 0 = clear drop, 1 = not a clear drop.
- **Testing experiments**
  - Classification errors: false positive (FP) and false negative (FN).
    - 100 images with precipitates/crystals: FP = 24%.
    - 100 images with clear drops: FN = 28%.
  - For our application, false positive errors are more critical than false negative since it would lead the operator to discard images with precipitate/crystals.
  - ⇨ The goal of the optimization of the neural network design was therefore to minimize FP errors.
  - Performance of the classification system:
    - Accuracy = (TP+TN)/(TP+TN+FP+FN)
    - Precision = TP/(TP+FP)
    - Recall = TP/(TP+FN)

[Accuracy Precision Recall] = [74% 0.73 0.76] obtained in this study
[Accuracy Precision Recall] = [88% 0.83 0.76] [3] single plate
[Accuracy Precision Recall] = [85% 0.24 0.66] [3] multiple plates

5

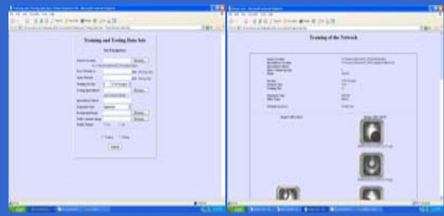
ACKNOWLEDGEMENTS

This project is part of the Northeast Structural Genomics Consortium (NESG) sponsored by the NIH for evaluating the feasibility, costs, economics of scale, and value of structural genomics, NIH P50 GM62413-05.

9

RESULTS

1. Remote Application Web-server Interface

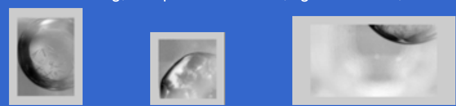


2. Classification with a New Image Database

- **Image Features for Neural Network**
  - Quantitative shape descriptions of the first-order histogram of Laplacian pyramid coefficients.
  - 8 statistics for each coefficient subset.
- **Training data set**
  - 500 images
  - 1/2 clear drops, 1/2 drops with precipitates/crystals
- **Testing experiments**
  - 100 images with precipitates/crystals: FP = 7%.
  - 100 images with clear drops: FN = 3%.

3. Examples of Misclassified Images

Motion blurring, complex boundaries, light reflection, etc.



Small bubble size, no sharp gradient, background influence, etc.

6

DISCUSSION

- Introduction of robotic manipulation of the crystals for HTP protein production requires the automation of image analysis of crystallization experiments for classification of solution content.
- The proposed feed forward neural network showed promising results in classifying microscopic images.
- Features of representation were computed from Laplacian pyramid expansion histograms. The histogram made the features invariant to orientation which was a desirable feature in order to be able to characterize the diversity and complexity of precipitate appearances.
- The Laplacian expansion provided a representation of the image edge and texture patterns at different scales with extremely fast implementation.

7

CONCLUSION

- Future applications of the project include further classification of the microscopic image database to separate crystals from precipitates and drops with organic matter.
- A parallel task of this project includes the creation of a web-based infrastructure for testing and development. An experimental test-bed for crystallization screening is currently being developed .

8