High-speed Recognition of Micro-array Genomic Imag Using Multi-scale Representations

Ya Wang, Elsa D. Angelini, David H. Kim, Andrew F. Laine

Heffner Biomedical Imaging Laboratory

Department of Biomedical Engineering, Columbia University, New York, 10027, NY, USA

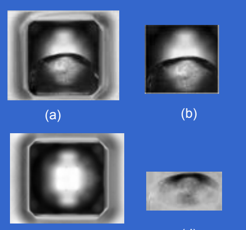
INTRODUCTION

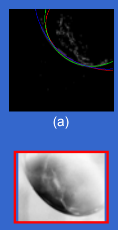
- This project is part of the Northeast Structural Genomics Consortium (NESG). The goal of this consortium is to develop efficient and integrated technologies for high-throughput (HTP) protein production and 3D structure determination.
- This project focuses on the design of an image analysis system to classify protein crystal structures in a production oriented environment.
- The method performs classification of microscopic images as clear droplets versus non-clear droplets (precipitates and crystals).
- Using expert classification for ground truth, current results show high classification accuracy with a large image datasets.

1

METHODOLOGY

1. Preprocessing of Microscopic Images



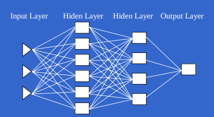


(a) Original image with an oil droplet containing precipitates. (b) Cropped image with Radon transform. (c) Background image acquired without drops. (d) Pre-processed image with Ellipsoidal Hough transform.

(a) Ellipsoidal Hough transform [1] to detect the three most probable ellipses, plotted over the edge map of the pre-filtered image. (b) Pre-processed 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



4. We used the quantitative shape descriptions of a first-order histogram combined with the power spectrum and autocorrelation information:

$$\text{Mean:}\quad S_M \equiv \overline{b} = \sum_{b=0}^{L-1} b\,P(b)$$

$$\text{Standard Deviation:}\quad S_D = \left[\sum_{b=0}^{L-1} (b-\overline{b})\right]^{1/2}$$

$$\text{Skewness:}\quad S_S = \frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}(b-\overline{b})^3 P(b)$$

$$\text{Kurtosis:}\quad S_K = \frac{1}{\sigma_b^4}\sum_{b=0}^{L-1}(b-\overline{b})^4 P(b) - 3$$

$$\text{Energy:}\quad S_N = \sum_{b=0}^{L-1}[P(b)]^2$$

$$\text{Entropy:}\quad S_E = -\sum_{b=0}^{L-1} P(b)\log_2[P(b)]$$

$$\text{Power:}\quad S_P = \sum^{L-1} fft(\mathrm{Im})^2$$

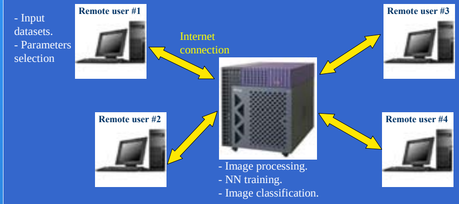
$$\text{Autocorrelation:}\quad S_A = \sum_{i=0}^{L-1}\sum_{t=0}^{T} \mathrm{Im}(i)\bullet\mathrm{Im}(t+i)$$

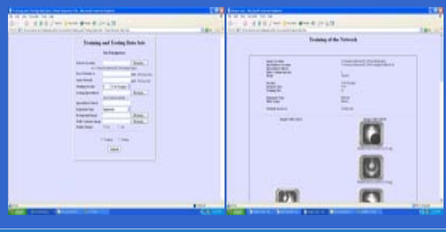
2

METHODOLOGY

5. Remote Application Server

Web page interface managed by a Matlab© Server.

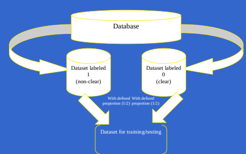




3

DATA

1. Microscopic images were acquired with a CCD camera under robotic control.
2. Gray scale 8-bits images saved in tiff format.
3. Image Database of 5,000 manually classified images:
   - 2500 drops containing precipitates and/or crystals.
   - 2500 clear drops (clear, skin, etc)



4

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.
- Most 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.

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.

9

RE TS

1. Classification with the Database

► **Image Features for Neu ork**

Quantitative shape descripti first-order Laplacian pyramid coefficients histogram con th the power spectrum and autocorrelation information

8 statistics for each Lapla ficient subset with totally 5 subsets. (8*5=40)

► **Binary classification**

0 = clear drop, 1 = not a c

► **Training data set**

100, 500, 1000, 1500, 2000 00, 3500 images

1/2 clear drops, 1/2 drops w tates/crystals

► **Testing experiments**

200 images with precipitate TP = 90%.

200 images with clear drops %.

► **Definition of the classifi stem**

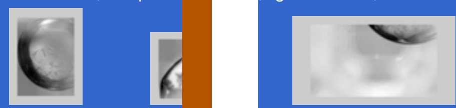
Accuracy = (TP+TN)/2 =

TP= True Possitive (Per = Correctly classified images with precipitates c / Total images with precipitates or crystals*100

TN = True Negative age) = Correctly classified images with clear drops/ ges with clear drops*100%

2. Examples of Misclass ages

Skin effects, complex b , light reflection, etc.



Small bubble size, no sharp background influence, etc.

6

CONC ON

- We are working on further c on of the microscopic image database to separate crystal cipitates.
- A parallel task of this project the creation of a web-based infrastructure for testing and ent. An experimental test-bed for crystallization screen ently under development.

7

REFER ES

[1] Angelini, E. D., Y. Wang, et al. (2004). C f Micro Array Genomic Images with Laplacian Pyramidal Filters and Neural Ne shop on Genomic Signal Processing and Statistics (GENSIPS), Bal and, USA.

[2] D. H. Ballard, "Generalizing the Hough t etect arbitrary shapes," *Pattern Recognition*, vol. 13, pp. 111-122, 1981.

[3] P. J. Burt and E. H. Adelson, "The Lapla as a compact image code," *IEEE Transactions on Communications*, vol. 31, 1983.

[4] C. A. Cumbaa, A. Lauricella, N. Fehrma R. Collins, J. Luft, G. DeTitta, and I. Jurisica, "Automatic classification of sub-m in-crystallization trials in 1536-well plates," *Acta Crystallographica Section D- stallography*, vol. 59, pp. 1619-1627, 2003.

[5] I. Jurisica, P. Rogers, J. I. Glasgow, R. J . Wolfley, J. R. Luft, and G. T. DeTitta, "Improving objectivity and scalabi crystallization," *IEEE Intelligent Systems*, vol. 16, pp. 26-34, 2001.

8