Neural Network Scoring of Spots in X-Gal and –leu Plates

K. Jafari-Khouzani[a,b], H. Soltanian-Zadeh[b,c], R.L. Finley Jr.[d], and F. Fotouhi[a]

[a]*Department of Computer Science, Wayne State University, Detroit, MI 48202, USA*
[b]*Radiology Image Analysis Lab., Henry Ford Health System, Detroit, MI 48202, USA*
[c]*Electrical and Computer Engineering Dept., University of Tehran, Tehran 14395, Iran*
[d]*Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA*

E-mails: kjafari@rad.hfh.edu, hamids@rad.hfh.edu, rfinley@genetics.wayne.edu, fotouhi@cs.wayne.edu

***Abstract* - We have developed an image analysis system for scoring yeast growth and color development in images of 96-well plates, a common format for high throughput assays. We use a segmentation method to locate the plates and spots. Color histogram and wavelet features are extracted respectively from spots of X-Gal and –leu plates. Two artificial neural networks are separately employed to score spots on each plate. The performance of the system is evaluated using a data set of 50 images. The data set was divided into 25 training and 25 testing images. Accuracies of 99.7% and 95.2% have been achieved for scoring the X-Gal and –leu plates respectively.**

I. INTRODUCTION

High throughput assays have become an important part of the discovery process in biological and biomedical research. A variety of high throughput technologies have been developed, for example, to discover how genes and their encoded proteins function together in normal and disease states [1, 2]. Many of these technologies use the baker's yeast, *Saccharomyces cerevisiae*, either as a model organism to study biological processes that yeast have in common with humans, or as an assay system in which to test the functions of genes and proteins from humans or other model organisms [3]. For example, one technology known as the "yeast two-hybrid system" uses yeast cells essentially as test tubes in which to assay interactions between specific proteins from other organisms [4, 5]. Other assays have been developed that use yeast cells to screen for drugs or new drug targets [6-8]. One common feature of these and other yeast-based assays is that they measure the ability of the yeast to grow or to change color on solid media. In the yeast two-hybrid system, for example, the protein-protein interaction results in activation of two different reporter genes. One reporter gene is required for growth and the other encodes an enzyme that converts a colorless compound to a color in the visible spectrum. Generally, digital photographs are taken of the solid media containing the yeast, and the level of yeast growth and the color of the yeast colony for each assay is scored manually. This is a tedious time-consuming process that is particularly rate limiting for high throughput screens that conduct hundreds to thousands of these assays in each experiment.

We have developed an image analysis system for scoring yeast growth and color development in images of 96-well plates, a common format for high throughput assays. The system is employed to analyze the results of a high throughput yeast two-hybrid screen. In this screen, yeast is spotted on two flat plates in the common 96-well format, with twelve columns and eight rows of spots (see Fig. 1). Each spot represents a separate assay for protein-protein interactions using a modified version of the yeast two-hybrid system [9, 10]. One of the plates lacks a nutrient such that the yeast will not grow unless they are expressing one of the *LEU2* reporter genes (the "–leu plates"). The other plate contains the chromogenic substrate "X-Gal", which will be converted to a blue compound in yeast that expresses the other reporter gene, *lacZ*. This configuration provides two independent assays for protein-protein interactions; the same 96 spots are placed on both the –leu plate and the X-Gal plate. After the yeast are given a chance to grow by incubation for 2-4 days, the plates are photographed. According to convention [10], the level of growth on the –leu plate is scored on a scale of 0-3, where 0 indicates no growth and 3 indicates maximum growth. On the X-Gal plate, the level of reporter activity is scored as the amount of blue color of the yeast on a scale of 0-5, where 0 is white and 5 is dark blue. Fig. 2 shows samples of spots on X-Gal and –leu plates with their scores. As shown, in –leu plate the amount of growth is determined by the intensity and texture of the spot, while in X-Gal plate, the color specifies the score.

Several studies have shown that the amount of growth on –leu plates and the amount of blue color on X-Gal plates correlates with the level of reporter activity and the general affinity of the protein-protein interaction being assayed [11]. This quantitative information promises to be particularly useful for developing computational approaches to interpret protein interaction data.

Similar works have been done for segmentation and scoring the spots in different applications [12]-[19]. An automatic method has been presented in [12] for identification of bacterial types. Other works [13]-[19] are on the microarray segmentation and processing.

The outline of this paper is as follows. In Section II we describe the segmentation technique to locate the spots in each plate. In Section III we present the proposed features to be extracted from each spot. The classification method is presented in Section IV. Experimental results are explained in section V. Conclusions are presented in Section VI.

0-7803-9187-X/05/$20.00 ©2005 IEEE.

II. SEGMENTATION

As shown in Fig. 1, the image contains two plates located in a dark background. The images are of size 1200×1792 pixels. The segmentation and localization of the spots of each plate can be accomplished in two steps. In the first step, the plates are separated from the background. In the second step, the spots are segmented from each plate. In the following subsections each step is explained.

A. Segmentation of Plates

To separate the plates from the background, the vertical projection of the gray-scaled image is first used to find the location of the upper edge of the plate (see Fig. 3). The mean value of the projection is used as the threshold to find this location. Since all the images are taken at the same situations, the plates have a constant size. Therefore, we can separate the upper and lower background as shown in Fig. 4. Similarly, using the horizontal projection, we remove the background in the left and right parts of the image (see Fig. 4). Two thresholds are calculated for finding the left and right lines of the plates. Two signals with the length of 200 (pixels) are taken from left and right sides of the projection and thresholded by their mean values to find the left and right limits.

The middle line of the resulting image divides the image into two plates. At this point we can check if the image is rotated. A subimage is taken from where the upper right corner of the left plate meets the upper left corner of the right plate (shown in Fig. 5). This subimage contains edges of the two plates. Using the Radon transform orientation estimation method proposed in [20] we estimate the orientation of the plates. If the image is rotated, we correct the orientation of the original image and then follow all the previous steps to segment the plates.

The next step is to use the projection of each plate to locate the spots. Fig. 6 shows the horizontal and vertical projections of a segmented plate. To locate the spots, we calculate the correlations of the vertical and horizontal projections with shifts of a symmetric sinusoidal shape signal as shown in Fig. 7. This signal is a template signal produced by $f(x)=|\sin(x)|$ and has 12 local maxima for the vertical projection and 8 local maxima for the horizontal projection. The period of the signal is equal to the distance between the centres of two adjacent spots. If $f_v(x)$ is the template signal and $p_v(x)$ is the vertical projection, then local maxima of

$$r_v(k) = \mathrm{corr}(p_v(x), f_v(x-k))$$

occur where some local maxima of $f_v(x-k)$ match some local maxima of $p_v(x)$ (see Fig. 7). To find the location where all the local maxima of $f_v(x)$ match all local maxima of $p_v(x)$, we calculate $r_v(k)$ for $-M \le k \le M$ where M is the location where the first local maximum of $f_v(x)$ occurs. By finding the local maxima of $r_v(k)$ in this interval, we get the possible locations of the centers of the spots using the template signal. As shown in Fig. 7, we usually have more than one local maximum, and the global maximum in this interval may not necessarily correspond to the maximum matching between the template signal and the spots locations. This is due to the rapid oscillations of the projection at the corners (see Fig. 6), which are created by the edges of the frame. To select the right local maximum, notice that the derivative of the projection is higher at the corners (see Fig. 8). Thus, the mean value of the derivative of the projection at the right and left corners, where these rapid oscillations are expected to occur, can be used to select the correct local maximum. For each local maximum and its expected location of oscillations, we calculate the mean derivative and choose the one with the maximum mean value. In this way, we eliminate the edges of the frame. The final result for X-Gal plate is depicted in Fig. 9. The -leu plate is segmented similarly.

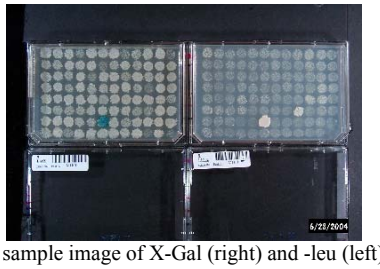
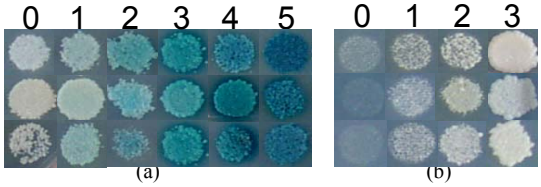


Fig. 1 A sample image of X-Gal (right) and -leu (left) plates.



(a) (b)

Fig. 2 Samples of spots on (a) X-Gal plate, and (b) –leu plate with their scores. As shown, in X-Gal plates the score depends on the color, and in the –leu plate the score depends on the intensity and texture of the spot.

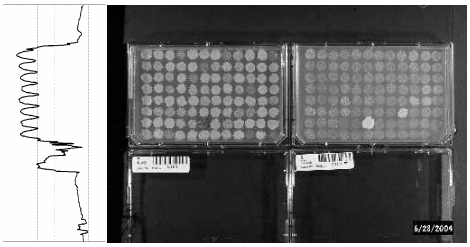


Fig. 3 Horizontal projection of the image to find the upper edge of the plates.

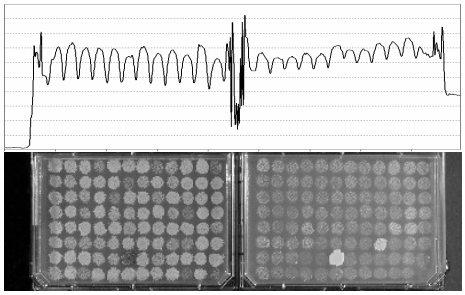


Fig. 4 Vertical projection of the image to find the left and right edges of the plates.

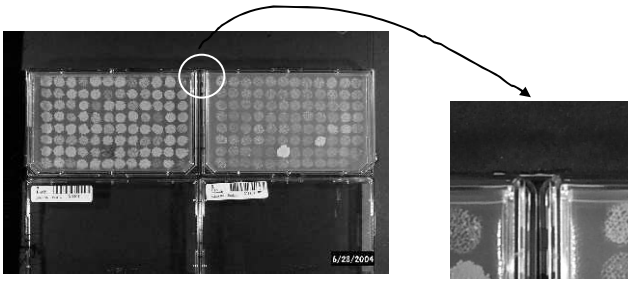


Fig. 5 Orientation adjustment is done using the specified location of the image. The Radon transform of this area can be used to estimate the orientation of the plates as described in [20].

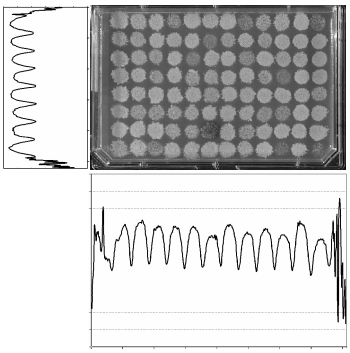


Fig. 6 The horizontal and vertical projections of the plate image to locate the spots.

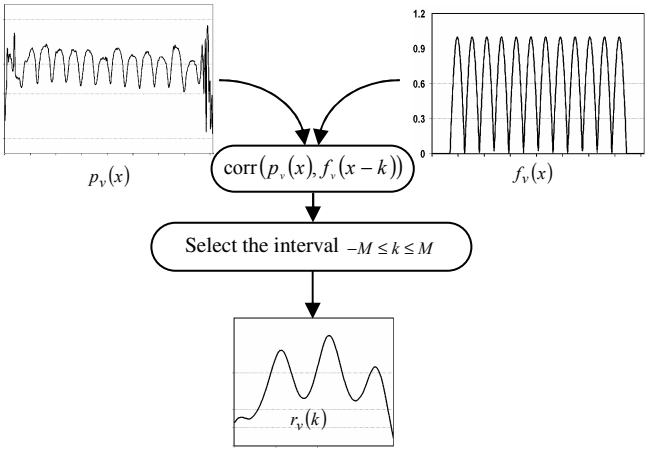


Fig. 7 The correlation of $f_v(x-k)$ and $p_v(x)$ for $-M \le x \le M$ where L is the size of $f_v(x)$, and M is the location where the first local maximum of $f_v(x)$ occurs.

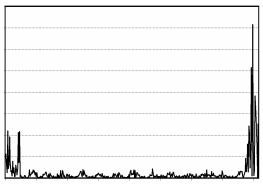


Fig. 8 The derivative of the vertical projection. As shown there are rapid oscillations at the left and right corners compared with the middle of the signal.

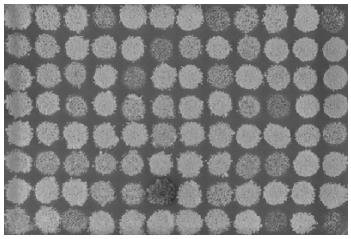


Fig. 9 The final segmentation result for X-Gal plate. The –leu plate is segmented similarly.

B. *Segmentation of Spots*

Following segmentation of the plates, we segment the spots. Like in Fig. 6, we calculate the horizontal and vertical projections of the segmented plates. Here, the rapid oscillations of the projections at the corners no longer exit. Therefore, after smoothing the projection signal we can locate the spots by calculating the local minima using the derivative of the signal. However, since the spots may not have uniform intensities, we may miss some local minima of the projection. Furthermore, due to the same problem, we may not have uniform spacing between the local minima. To avoid this problem, we construct the following sequence (see Fig. 10(a)):

$$f[n] = \begin{cases} 1 & n = \arg(\text{Local Min}(\text{projection})) \\ 0 & o.w. \end{cases}$$

and calculate its correlation with shifts of a signal shown in Fig. 10(b). This signal has the same size as $f[n]$ and is created by repeated and equally spaced Gaussian functions with the local maxima of 1. The distance between each two local maxima is equal to the distance between the centers of the spots. The number of the Gaussian functions is equal to the number of expected grid lines in each direction (horizontal or vertical). The signal shown in Fig. 10(b) is suitable for horizontal projection. By finding the maximum of the resulting correlation, we locate where these two signals have maximum matching and place the grid lines accordingly.

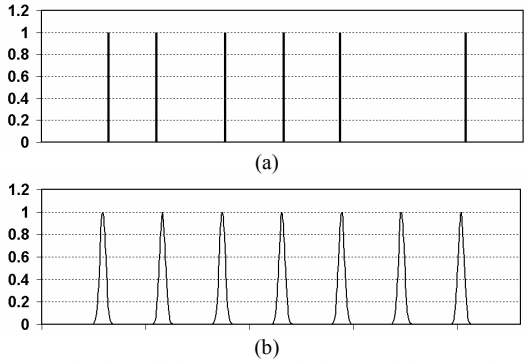


Fig. 10 Locating the grid lines. (a) Local minima of the projections. (b) a signal to locate accurately the grid lines. After the calculating the correlation of (a) with shifts of (b), the local maximum of the result is located in a suitable interval, to locate the grid lines.

By finding the separating lines as shown in Fig. 11, we locate the spots. In each block in Fig. 11, we calculate the

center of mass of the spot and center a circle around it. We can then use the intensity information inside the spot for feature extraction. The final segmentation of the spots in X-Gal and –leu plates is presented in Fig. 12.

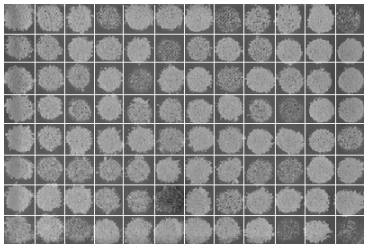
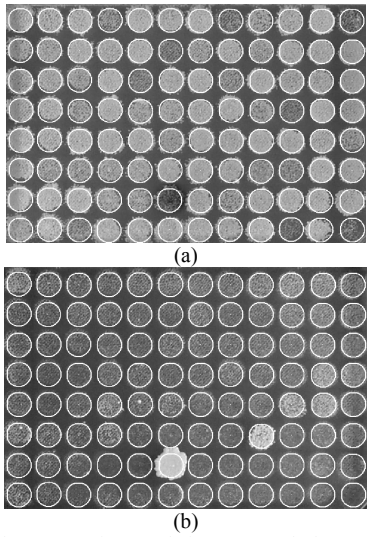


Fig. 11 Overlaying the grid lines on the X-Gal plate.



(a)

(b)

Fig. 12 The final segmentation result for (a) X-Gal plate, and (b) –leu plate.

III. Feature Extraction

To score the spots, we extract features from each spot and classify them using Artificial Neural Network (ANN). As mentioned earlier, in X-Gal plates the color determines the score, while in -leu plates, intensity variations specify the score. Therefore, we employ separate techniques to extract features from spots of each plate.

A. X-Gal Plates

Since in the X-Gal plates the score depends on the color variations, we use the color histogram of the image as a feature vector. We map the color image into a predefined and fixed color map to calculate the histogram of the colors. However, in practice, the color map may be large (210 in our experiments), so we need to reduce the size of the histogram. Each point in the color map is assigned an integer number between 0 and 5. The number is equal to the score in which that specific color occurs more frequently compared with the other scores. Therefore, we get a feature vector (histogram) of size 6 for X-Gal spots.

B. -leu Plates

Since in the -leu plates, the score depends on the intensity variations of the spot, it may depend on the local illumination. Therefore, we need to first normalize the intensities of the spots. Mean value of the spot is not a suitable value for normalization, since the growth area (the bright areas of the spot) may vary from spot to spot. To find a proper value, we first remove the background pixels from the spot and find the mean intensity of the rest. The resulting value can be used to normalize all pixel intensities before feature extraction. To find the background pixels, we map the color image into a predefined and fixed color map, which creates a color histogram. From our previous knowledge, we know which colors belong to background. Therefore, the background pixels are specified.

After the normalization, we calculate the wavelet features. We calculate the ordinary wavelet transform of each spot up to three levels using Daubechies wavelet basis with length 6 (db6). The resulting wavelet transform has 10 subbands. From each subband I, we calculate the following feature:

$$e_1^{(l,s)} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I_{i,j}^{(l,s)} \right|$$

where M and N are the dimensions of each subband, l is the decomposition level, and s represents the subband number within the decomposition level l. Therefore, 10 features are created in this way.

IV. Classification

To classify each spot into appropriate class, we use ANN. For X-Gal spots, we use a perceptron with 3 layers and 3, 3, and 1 neurons in the first, second, and third layers, respectively. The first two layers use sigmoid transfer functions; while the last layer use a saturated linear transfer function. Since the output of the network is in the interval [0,1], we multiply it by 5 to get the possible values in the range [0,5]. To find the score of a spot, the features are calculated and fed to the neural network and the output of the network is calculated and multiplied by 5, then rounded to the nearest integer. Therefore, the final output values are $\{0,1,2,3,4,5\}$.

For -leu spots, we use similar ANN with the difference that the output is multiplied by 3. Thus, the output values are $\{0,1,2,3\}$. The two networks are trained separately.

V. Experimental Results

To evaluate the performance of the system, we used a data set of 50 images. The segmentations of the spots were done correctly (with manual inspection). We used 25 images for training and 25 for testing. All the spots were also scored manually by an expert to evaluate the system. After training the neural network for scoring the X-Gal plates, we obtained 98.3% accuracy for training samples. Note that we compared the automatic scores with the manual scores. The manual scores might not be consistent. Therefore, the real accuracy could be higher. Using the trained neural network for the 25 test images, we obtained an accuracy of 99.7%. The reason we get higher accuracy for testing set is because for training set

we chose the plates with more number of green and blue spots (scores 1-5), in order to have the best training set (in practice most of the spots are of score zero). Therefore the training set is more difficult to score than the testing set.

After training the neural network for scoring the -leu plates, we obtained 95.2% accuracy for training samples. Using the trained neural network for the 25 test images, an accuracy of 94.7% was achieved.

VI. Conclusion

An automatic method was presented to segment and score the spots of X-Gal and –leu plates. Color histogram and wavelet features were extracted respectively from spots of X-Gal and –leu plates. Neural networks were used for scoring the spots using the extracted features. Experimental results showed reliability of the system for scoring the spots.

Acknowledgment

We would like to thank Jodi Parrish and Barbara Basberg for helping with data acquisition and system evaluation.

References

[1] E. Phizicky, P.I. Bastiaens, H. Zhu, M. Snyder, and S. Fields, "Protein analysis on a proteomic scale," *Nature*, vol. 422(6928), pp. 208-215, 2003.

[2] S. Fields, Y. Kohara, D.J. Lockhart, "Functional genomics," *Proc Natl Acad Sci U S A* 1999, vol. 96(16), pp.8825-8826.

[3] T.R. Hughes, M.D. Robinson, N. Mitsakakis, and M. Johnston, "The promise of functional genomics: completing the encyclopedia of a cell," *Curr Opin Microbiol*, vol. 7, no. 5, pp. 546-554, 2004.

[4] R. Brent and R.L. Finley Jr., "Understanding gene and allele function with two-hybrid methods," *Annu Rev Genet*, vol. 31, pp. 663-704, 1997.

[5] S. Fields and O. Song, "A novel genetic system to detect protein-protein interactions," *Nature*, vol. 340(6230), pp. 245-246, 1989.

[6] A. Weiss, J. Delproposto, and C.N. Giroux, "High-throughput phenotypic profiling of gene-environment interactions by quantitative growth curve analysis in Saccharomyces cerevisiae," *Anal Biochem*, vol. 327, no. 1, pp. 23-34, 2004.

[7] T.R. Hughes, "Yeast and drug discovery" *Funct Integr Genomics*, vol. **2**(4-5), pp. 199-211, 2002.

[8] R.A. Butcher, and S.L. Schreiber, "Identification of Ald6p as the target of a class of small-molecule suppressors of FK506 and their use in network dissection," *Proc Natl Acad Sci U S A* 2004, vol. 101, no. 21, pp. 7868-7873.

[9] C.A. Stanyon, G. Liu, B.A. Mangiola, N. Patel, L. Giot, B. Kuang, H. Zhang, J. Zhong, and R.L. Finley Jr., "A Drosophila protein-interaction map centered on cell-cycle regulators," *Genome Biol*, vol. 5(12), pp. R96, 2004.

[10] J. Zhong, H. Zhang, C.A. Stanyon, G. Tromp, and R.L. Finley Jr., "A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating," *Genome Res*, vol. 13(12), pp. 2691-2699, 2003.

[11] J. Estojak, R. Brent, and E.A. Golemis, "Correlation of two-hybrid affinity data with in vitro measurements," *Mol Cell Biol*, vol. 15(10), pp. 5820-5829, 1995.

[12] S. Trattner, H. Greenspan, G. Tepper, and S. Abboud, "Automatic identification of bacterial types using statistical imaging methods," *IEEE Trans. Medical Imaging*, vol. 23, no. 7, pp. 807-820, July 2004.

[13] R. Nagarajan and C.A. Peterson, "Identifying spots in microarray images," *IEEE Trans. Nanobioscience*, vol. 1, no. 2, pp. 78-84, June 2002.

[14] M. Katzer, F. Kummert, and G. Sagerer, "Methods for automatic microarray image segmentation," *IEEE Trans. NanoBioscience*, vol. 2, no. 4, pp. 202-214, Dec. 2003.

[15] R. Nagarajan, "Intensity-based segmentation of microarray images," *IEEE Trans. Medical Imaging*, vol. 22, no. 7, pp. 882-889, July 2003.

[16] A.W.-C. Liew, H. Yan, and M. Yang "Robust adaptive spot segmentation of DNA microarray images," *Pattern Recognition*, vol. 36, no. 5, pp. 1251-1254, May 2003.

[17] R. Hirata Jr., J. Barrera, R.F. Hashimoto, D.O. Dantas, and G.H. Esteves, "Segmentation of microarray images by mathematical morphology," *Real-Time Imaging*, vol. 8, no. 6, pp. 491-505, Dec. 2002.

[18] X. Wang, N. Jiang, X. Feng, Y. Xie, P.J. Tonellato, S. Ghosh, and M.J. Hessner, "A novel approach for high-quality microarray processing using third-dye array visualization technology," *IEEE Trans. NanoBioscience*, vol. 2, no. 4, pp. 193-201, Dec. 2003.

[19] A. P.G. Damiance Jr., L. Zhao, and A. C.P.L.F. Carvalho, "A dynamical model with adaptive pixel moving for microarray images segmentation," *Real-Time Imaging*, vol. 10, no. 4, pp. 189-195, August, 2004.

[20] K. Jafari-Khouzani, H. Soltanian-Zadeh, "Radon Transform Orientation Estimation for Rotation Invariant Texture Analysis", *IEEE Trans. Pattern Analysis & Machine Intelligence*, in press, 2005.