

Automatic microarray gridding by mathematical morphology

DANIEL OLIVEIRA DANTAS and JUNIOR BARRERA

Universidade de São Paulo (USP), Brazil
{ddantas,jb}@ime.usp.br

1. Introduction

DNA chips (i.e., microarrays) biotechnology [2,3] is a hybridization (i.e., matching of pairs of DNA) based process that makes possible to quantify the relative abundance of mRNA from two distinct samples by analyzing their fluorescence signals. This technique requires robotic placement (i.e., spotting) of thousands of cDNAs (i.e., complementary DNA) in an array format on glass microscope slides. The spotted cDNAs are the hybridization targets for the mRNA samples.

The two different samples of mRNA, usually labeled with Cy3 and Cy5 fluorochromes, are cohybridized onto each spotted gene. After hybridization, one digital image is acquired for each fluorochrome wavelength. Then, it is necessary to recognize each gene, by its position in the array, and to estimate its signal (i.e., hybridization information). For that, it is necessary to segment the image in three classes of objects: subarrays (i.e., set of grouped spots), spot box (i.e., the rectangular neighborhood that contains a spot) and spot (i.e., region of the image where there exists signal). In this paper we present an improvement in the subarray gridding step published as “Segmentation of Microarray Images by Mathematical Morphology” [1].

2. Subarray gridding

Our objective is to draw a grid onto the microarray images, that correctly identifies the regions of interest (ROI) of each subarray, with minimal user interaction. The algorithm requires some parameter settings that can be reused for a whole set of microarrays spotted with the same geometrical characteristics and scanned with the same resolution. The parameters are the number of rows and columns of subarrays, number of rows and columns of spots per subarray, vertical and horizontal distance between spots centers, vertical and horizontal distance between subarrays, and spot diameter.

The procedure of subarray gridding is composed of the two following steps: morphological filtering and gridding correction. Depending on case, two preliminary steps may or may not be also required. A registra-

tion step is required if the image channels are not registered. A correction of rotation step is required if the spots are not aligned with the image borders.

3. Morphological filtering

Usually the experiment gives us an image with two channels, one for Cy3 and other for Cy5. Our system uses a composition, say f , of both channels to extract the grid information. In our tests, we used either union or sum of the two channels and the results were equivalent.

The morphological filtering is not applied in the channels composition called f , but in the projection profiles of the image f . The projection profiles are defined as the sum of the pixel values in each pixel row or pixel column. That gives us two unidimensional signals. The horizontal profile is given by the sum of pixels in each row and the vertical is given by the sum of pixels in each column. The idea behind this procedure is that the signal segmentation problem is a good approximation to the image segmentation problem and much simpler to solve.

The first filter tries to group together the vertical profiles of spots that belong to the same column of subarrays, and, respectively, horizontal profiles of spots that belong to the same row of subarrays. This filter is a morphological closing. The size of the structuring element is chosen based on the diameter of the spot and on the distances between subarrays in pixels.

After the first filter we can notice better the groups of peaks between the deep valleys where the spaces between subarrays are supposed to be located. We would like to eliminate the narrow peaks keeping only the larger ones corresponding to the groups of spot profiles that belong to the same column or row of subarrays. This is done by a morphological opening.

In the last filter we calculate the negation of regional minima to find the regions inside subarrays. Now we have a binary signal valued nonzero in the regions supposed to be subarrays, and zero in the regions supposed to be between subarrays. The next step, found to be useful in some noisy images, is to eliminate the connected components that touch the borders, that is, if a subarray group is found to be including the first or the last point of the signal, it is eliminated. This is done because, in some noisy images, the border noise can be identified as a region belonging to a subarray.

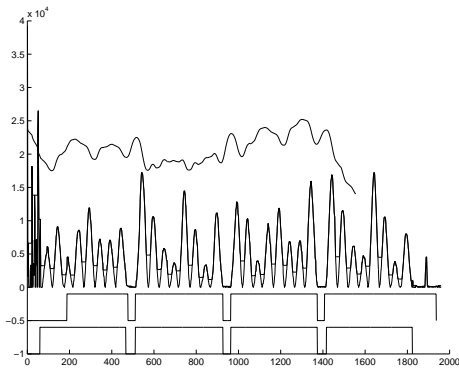


Figure 1. Vertical profile gridding correction.

At this point we have signals in which the nonzero pixels show the regions inside the subarrays.

4. Gridding correction

Usually, after the Morphological Filtering step, good images, that is, images with few noise, small rotation and with subarrays well aligned, give a very good subarray gridding result. But bad images that give wrong results occur often so we added a Gridding Correction step based on the slide geometry.

This step runs when the difference between some estimated subarray size or distance and the theoretical subarray size or distance, given by the user, is bigger than a certain tolerance. In our algorithm, the tolerance used is half a spot distance.

The final subarray position is such that, satisfying the tolerance, maximizes the moving average of the projection profiles. The moving average is taken from regions with the theoretical subarray size. We expect that a misaligned subarray grid will have a low average profile because it includes a region between subarrays.

Figure 1 shows four superimposed plots. From top to bottom, moving average of profile in a region with size equal to one theoretical subarray size with right extremity in that point; image profile; subarrays positions estimated by the morphological filter; subarrays positions after the Gridding Correction step.

The algorithm uses the subarrays positions that satisfy the tolerance as reference. It also uses the profile and geometry information. When few subarrays are considered correct, the algorithm uses those closest to the center of the image as reference, because this region is the least prone to noise.

When no subarray is considered correct, we re-

place all grids. The new grid is found by an optimization procedure that maximizes the subarrays moving averages. The subarray size considered is the one given by the user. The location of the first grid and the distances, all considered equal, are such that the sum of the subarrays moving average is maximum.

5. Conclusion

The proposed subarray gridding technique improves the robustness of the subarray gridding process and correctly set the grids of noisy images. It has three main steps: parameter setting, morphological filtering and gridding correction.

The gridding solution proposed is implemented under MATLAB, using the MMORPH toolbox (<http://www.mmorph.com/>) for mathematical morphology.

User interaction is required only to check the final results and in the first step, the parameter settings. The parameters can be saved to later use with a whole family with the same geometric and scanning configurations.

The technique was tested with a variety of images from different microarray spotter and scanners. Some mistakes occur when subarrays are not well aligned. It happens when the spotter needle is misaligned. Rotated images grid correctly when the user goes through the rotation correction step.

A future step of this research could be to try to segment subarrays individually. This would solve the misaligned needle problem and, probably, images with a small rotation degree would not need the rotation correction step.

References

- [1] Roberto Hirata Jr., Junior Barrera, Ronaldo Fumio Hashimoto, Daniel Oliveira Dantas, and Gustavo Henrique Esteves, *Segmentation of Microarray Images by Mathematical Morphology*, *Real-Time Imaging* **8** (2002), no. 6, 491-505.
- [2] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*, *Science* **270** (1995), 467-470.
- [3] M. Schena, R. A. Heller, T. P. Theriault, K. Konrad, E. Lachenmeier, and R. W. Davis, *Microarrays: biotechnology's discovery platform for functional genomics*, *Trends in Biotechnology* **16** (1998), no. 7, 301-306.