Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

# Enhancing the quality metric of protein microarray image*

WANG Li-qiang (王立强)†, NI Xu-xiang (倪旭翔), LU Zu-kang (陆祖康),
ZHENG Xu-feng (郑旭峰), LI Ying-sheng (李映笙)

(*State Key Lab of Modern Optical Instrumentation*, *Zhejiang University*, *Hangzhou 310027*, *China*)

†E-mail: wangliqiang@zju.edu.cn

Received July 21, 2003; revision accepted Sept. 21, 2003

**Abstract:** The novel method of improving the quality metric of protein microarray image presented in this paper reduces impulse noise by using an adaptive median filter that employs the switching scheme based on local statistics characters; and achieves the impulse detection by using the difference between the standard deviation of the pixels within the filter window and the current pixel of concern. It also uses a top-hat filter to correct the background variation. In order to decrease time consumption, the top-hat filter core is cross structure. The experimental results showed that, for a protein microarray image contaminated by impulse noise and with slow background variation, the new method can significantly increase the signal-to-noise ratio, correct the trends in the background, and enhance the flatness of the background and the consistency of the signal intensity.

**Key words:** Protein microarray, Image enhancement, Filter, Noise
**doi:**10.1631/jzus.2004.1621 **Document code:** A **CLC number:** TN911.73

## INTRODUCTION

The development of protein microarrays has accelerated within the past few years. Protein microarray can be widely used in diagnostics, drug screening and testing, disease monitoring, drug discovery and medical research (Jocelyn and Leodevico, 2002). A protein microarray is a small surface–for example, a microscope slide–onto which the protein targets are immobilized, then hybridized and labeled by fluorescent probes. To analyze the microarray, a confocal microarray reader uses a laser to excite the fluorophores and a photomultiplier tube (PMT) to detect the resulting fluorescence; and the system produces an image that shows the intensity of the fluorophores. Typically, a protein microarray image contains hundreds to thousands of target spots varying from tens to hundreds of microns in diameter. Fig.1 shows a part (one block, total 10 blocks) of an HCV (hepatitis C virus) protein microarray image obtained by our developing confocal microarray scanner in 10 μm resolution. The diameter of each spot is about 150 μm; the image is 16 bits TIFF format. For a 25 mm×75 mm protein microarray scanned at 10 μm resolution and digitized by a 16 bits A/D converter, the image size will be nearly 36 M bytes.

Due to the huge amount of data, automatic analysis for protein microarray images has become indispensable (Yang *et al*., 2002). However, this has proved to be difficult due to the poor signal-to-noise ratio (SNR), impulse noise produced by the photomultiplier tube, large intensity variation within spots and uneven background.

In order to eliminate the above problems and

* Project (No. 30227002) supported by the National Natural Science Foundation of China

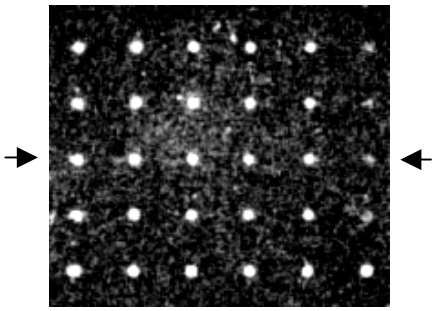


**Fig.1 One block of an HCV protein microarray image obtained by our confocal microarray reader**

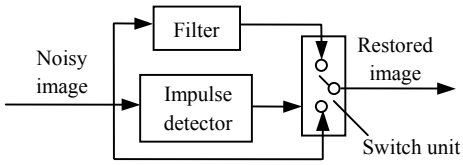



**Fig.2 A general framework of switching scheme-based image filter**

improve the quality metric of the protein microarray image, we use an adapted switching median filter to reduce the impulsive noise and a top-hat filter to correct the trends in background.

## NOISE REDUCTION

The various types of noise in the protein microarry image can be divided into two categories: source noise and detector noise. Examples of source noise are photon noise and dust on the slides. For a PMT, detector noise includes dark current noise and shot noise.

The background derived from the surface fluorescence by laser excitation is usually governed by the Poisson process, which can be approximated by a normal distribution when the arrival rate, or the accumulation of photons, is large enough. This property can be readily assessed by the histogram of any background region of the microarray image. Therefore, background noise can be simulated by a normal distribution (Balagurunathan and Dougherty, 2002). The target spots can also be modeled as bell-shaped Gauss distribution (Chen *et al.*, 2002).

To avoid the damage of good pixels in the protein microarray image, we use a median-based filter with switching scheme. Fig.2 shows a general framework for this kind of algorithm (Wang and Zhang, 1999). It detects whether the current pixel is contaminated at each pixel location. Then, for the corrupted pixels, median filtering is activated, while the noise-free pixels are left unaltered. Since not every pixel is filtered, undue distortion can be avoided. The impulse detection becomes therefore crucial to the subsequent filtering.

Considering the Gauss distribution of the intensity of both background and target spots in the protein microarray image, we propose a novel adaptive median filter that employs the switching scheme based on local statistics characters, which realizes the impulse detection by using the difference between the standard deviation of the pixels within the filter window and the current pixel of concern.

Consider a window $W$ defined in terms of the image coordinates symmetrically surrounding the current or original pixel (Chen and Wu, 2001).

$$W=\{(s,t)|\ -h<s<h,\ -h<t<h\} \tag{1}$$

where $h$ denotes the window size. The output of median filter applied to the origin pixel $X_{ij}$ can be described as

$$Y_{ij}=median(\{X_{i-s,\ j-t}|(s,t)\in W\}) \tag{2}$$

For current pixel $X_{ij}$ under consideration, we first define the difference $d_{ij}$ as

$$d_{ij}=|Y_{ij}-X_{ij}| \tag{3}$$

The objective of the impulse detector is to determine whether the current pixel is corrupted. The decision-making mechanism is realized by employing an adaptive threshold $T_{ij}$

$$T_{ij}=\overline{X}_{ij}+k\sigma_{ij} \tag{4}$$

where $k$ is the threshold adjustment parameter, $\overline{X}_{ij}$

is the average value of pixels in window $W$ surrounding the current pixel $X_{ij}$,

$$\bar{X}_{ij} = \frac{1}{(2h+1)^2} \sum \{X_{i-s,j-t} \mid (s,t) \in W\} \tag{5}$$

$\sigma_{ij}$ is the standard deviation of the pixels in window $W$ surrounding the current pixel $X_{ij}$,

$$\sigma_{ij} = \sqrt{\frac{1}{(2h+1)^2 - 1} \sum_{(s,t \in W)} (X_{i-s,j-t} - \bar{X}_{ij})^2} \tag{6}$$

For a pixel $X_{ij}$, if $d_{ij} \geq T_{ij}$, it is regarded as an impulse, otherwise, the impulse detector assumes $X_{ij}$ is noise-free. So the final output of the proposed filter is:

$$\hat{X}_{ij} = \begin{cases} Y_{ij} & \text{if } (d_{ij} \geq T_{ij}) \\ X_{ij} & \text{otherwise} \end{cases} \tag{7}$$

The window size $h$ and threshold adjustment parameter $k$ will influence the performance of the proposed filter. In our experiments, $h$=1,2,3, corresponding to 3×3, 5×5, 7×7 window respectively and $k$=0.5, 1, 1.5, 2 were applied. We observed that when $h$=2 and $k$=1, the proposed filter yielded satisfactory result in removing the impulse noise of the image in Fig.1. The resulting image is shown in Fig.3. The intensity plots of the third row in Fig.1 and Fig.3 (marked by arrow) are shown in Fig.4. We can see that the impulsive noise both in target spots and background is eliminated very well.

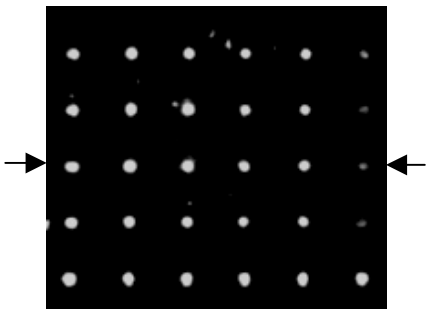


**Fig.3 Protein microarray image after filtered by the proposed filter, where $h$=2 and $k$=1**

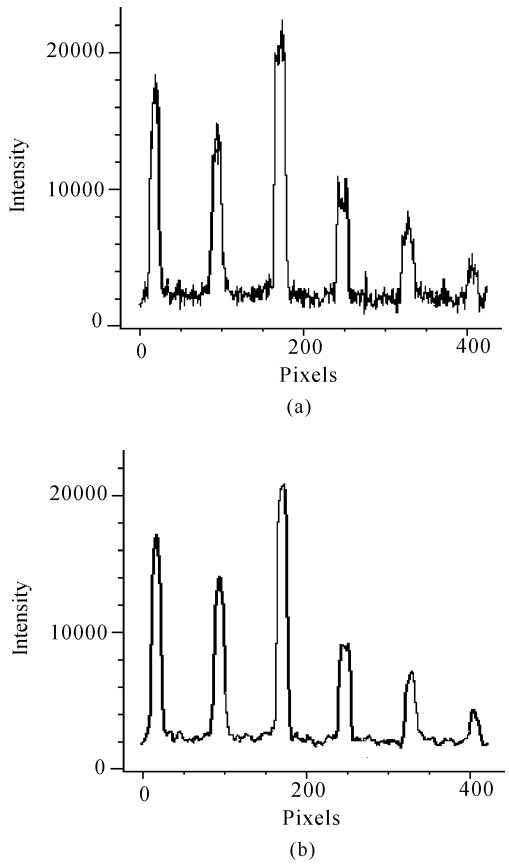



**Fig.4 Intensity plots of the third row of the images**
(a) vs Fig.1; (b) vs Fig.3, marked by arrow

## BACKGROUND VARIATION CORRECTION

Another problem with protein microarray image is that the brightness in the background varies; this may be caused by large area stained fluorescent probes or warp of the microarray slides, or other factors. Background variation makes comparison of similar features in different parts of an image difficult. It can be estimated using a morphological opening and removed by a top-hat filter (Yang *et al*., 2002; Glasbey and Ghazal, 2003).

Let $S$ be a structuring element which interacts with the protein microarray image.

$$S=\{(m, n)| -L<m<L, -L<n<L\} \tag{8}$$

where $L$ denotes the size of the structuring element; a morphological erosion operator $\varepsilon$ and dilation $\delta$ of

a digital image $X$ with the structuring element $S$ being defined as:

$$\varepsilon_S(X) = \min\{X_{i+m,j+n} | (m,n) \in S\} \quad (9)$$

$$\delta_S(X) = \max\{X_{i-m,j-n} | (m,n) \in S\} \quad (10)$$

Erosion replaces the value of the image $X$ at pixel $X_{ij}$ by the minimum of the values in the image region determined by $S$, whose center is $X_{ij}$ and size is $(2L+1)\times(2L+1)$; whereas dilation replaces $X_{ij}$ by the maximum of the values in the same image region. A morphological operator opening $\gamma$ is defined as

$$\gamma_S(X) = \delta_S(\varepsilon_S(X)) \quad (11)$$

So using a structuring element which is larger than the target spots, first we replace each pixel by the minimum local intensity in the region and then perform a similar operation on the resulting image, using the local maximum, all the target spots will be removed and the background for the entire slide will be generated. Through a top-hat filter $T$, which is defined as,

$$T_S(X) = X - \gamma_S(X) \quad (12)$$

the background trend will be corrected and the target spots will become more distinct.

Because the diameters of the protein microarray spots in our experiment were about 150 μm and the scanning resolution of our confocal microarray reader was 10 μm, the diameter of each spot was about 15 pixels. We found that cross structure can save much processing time with only a little quality loss, so we choose 17×17, 21×21, and 25×25 cross structuring elements for top-hat filtering, as shown in Fig.5. The 21×21 cross structure is a better tradeoff because it is more steady than 17×17 and faster than 25×25. In Fig.6, (a) is the background of the image in Fig.3 obtained by morphological opening using a 21×21 cross structure, (b) is the top-hat filtering result of the image in Fig.3 by a 21×21 cross structure. The intensity plots of the third row of the images (marked by arrow) are shown in Fig.7.

## QUALITY METRIC EVALUATION

Considering important factors that affect the measurement quality of protein microarray image, we propose a quality metric, which includes signal-to-noise ratio (SNR) of each spot, flatness of background and consistency of signal intensity. Quality metric was discussed in some papers (Brown *et al*., 2001; Chen *et al*., 2002; Wang *et al*., 2001), but they did not consider the influence of enhancing algorithm for microarray image. Here we evaluate the presented enhancing algorithm from the three aspects stated above.

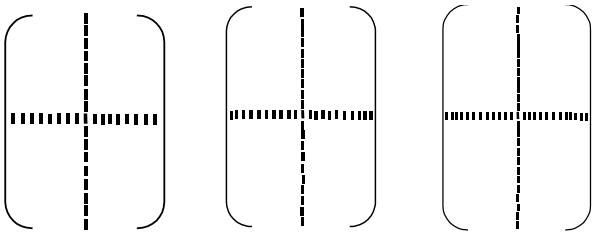
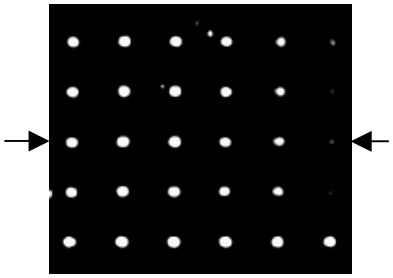


**Fig.5 Cross structuring elements for top-hat filtering. They are 17×17, 21×21 and 25×25 from left to right respectively**

(a)



(b)

**Fig.6 (a) The background image obtained by morphological opening using a 21×21 cross structuring element; (b) The top-hat filtering result after subtracting the background trend**

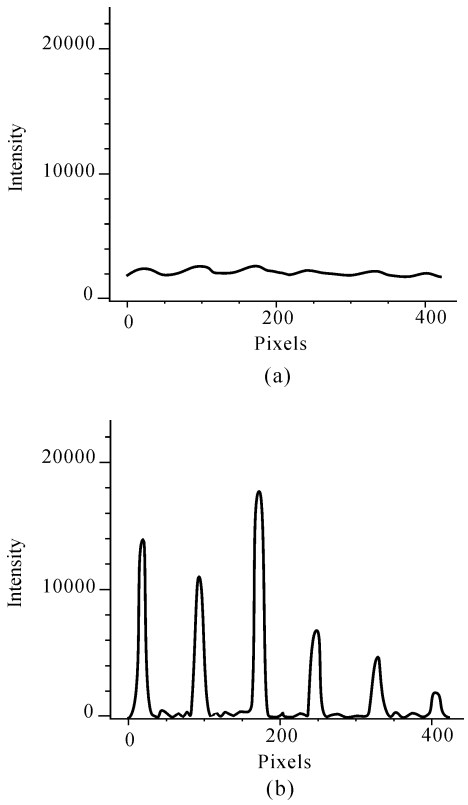



**Fig.7 Intensity plots of the third row of the images in Fig.6**
(a) vs Fig.6a; (b) vs Fig.6b

## Signal-to-noise ratio

The SNR quantifies how well one can resolve a true signal from the noise of the system. It is clear that when the SNR is low, the intrinsic variation in the data is high and confidence in the accuracy of the data is low. For the protein microarray image, we define the SNR of a target spot as:

$$SNR = \frac{S_p - B_p}{\sigma_p} \tag{13}$$

$S_p$ is the mean intensity of the target spot; $B_p$ is the mean intensity of the local background; $\sigma_p$ is the standard deviation of the local background. A higher SNR indicates higher signal over background noise, a SNR of 3 is commonly considered the lower limit for accurate detection. Signal may be detected below this value, but the accuracy of quantitative measurements decreases significantly. From Eq.(13), the SNR can be maximized by increasing signal, decreasing background, or decreasing noise (i.e. standard deviation of the local background).

It is clear that we decrease the noise by median-based filter with the switching scheme presented above, and we decrease background by top-hat filter. Fig.8 shows the SNR of each target spot in microarray image; $S_p$, $B_p$ and $\sigma_p$ are calculated using GenePix4.1 software. The pixels used for computing $S_p$ and the pixels used for computing $B_p$ and $\sigma_p$ are shown in Fig.9. The details can be found in (Axon, 2002). Evidently, the enhancing algorithm presented above can improve the SNR of the protein microarray image significantly.

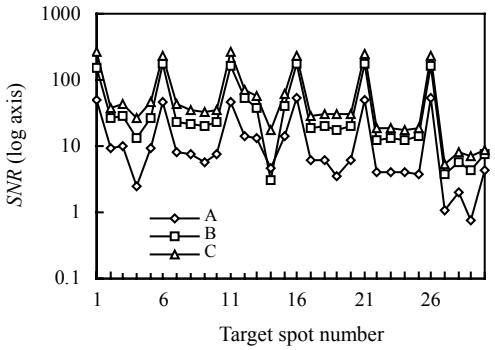



**Fig.8 SNR of each target spot in protein microarray image**
A: original image (Fig.1); B: median-based filtered image (Fig.3); C: top-hat filtered image (Fig.6b)

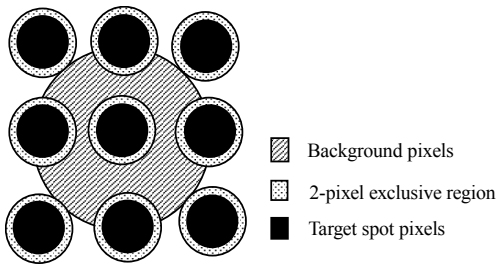



**Fig.9 Pixels of the target spot and the pixels of the background in GenePix4.1 software**

## Flatness of background

In protein microarrays, the primary source of background signal is non-specific hybridization in the same plane of focus as the target spots; other

sources also include biochemical factors such as probe purity and hybridization uniformity, as well as instrument factors such as stray photons and electronic noise inherent in all PMTs. Background abnormality can cause problems for signal detection or incorrect measurement of the local background level, so the flatness of the background signal reflects these factors; the more uniform the intensity of background, the higher the quality of the protein microarray image.

Fig.10 plots the local background intensity $B_p$ of each spot in different protein microarray images (A vs Fig.1, B vs Fig.3 and C vs Fig.6). From Fig.10, we can see that the background fluctuation is almost unchanged by median-based filter we presented, but after top-hat filtering, the background not only decreases evidently, it becomes more uniform.

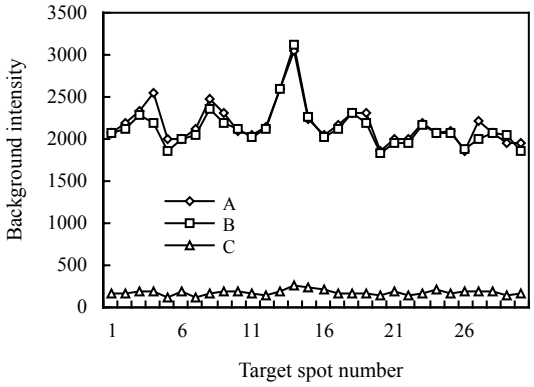



**Fig.10 Background intensity of each target spot in protein microarray images**
A: original image (Fig.1); B: median-based filtered image (Fig.3); C: top-hat filtered image (Fig.6b)

## Consistency of signal intensity

Since the intensity profile of the target spot is bell-shaped, just like Gaussian distribution, we define the consistency of signal intensity as the coefficient of variation (CV) of the target spot's intensity. In cases where contamination crosses the target spot or strong impulsive noise sits atop the target area, the detected signal of the target spot may not truly reflect the actual signal, so its standard deviation will be unusually high. We use the coefficient of variation to normalize the standard deviation relative to the mean signal intensity of the target spot:

$$CV_p = \frac{\sigma_s}{S_p} \tag{14}$$

$\sigma_s$ is the standard deviation of the target spot's intensity.

Fig.11 shows the coefficients of variation $CV_p$ of all spots in different protein microarray images (A vs Fig.1, B vs Fig.3 and C vs Fig.6). From Fig.11, we can see that although both B and C are smaller than A, C (after top-hat filtering) is a little bigger than B (just after median-based filtering). This is caused by the property of the top-hat filter; it eliminates the entire background trend; and at the same time, also decreases the mean intensity of the target spot $S_p$. This property can be seen clearly from Fig.7. However, the standard deviation of the target spot's intensity $\sigma$ is decreased a little due to the flatness of the morphological opening, so the $CV_p$ of all spots after top-hat filtering will be increased a little.

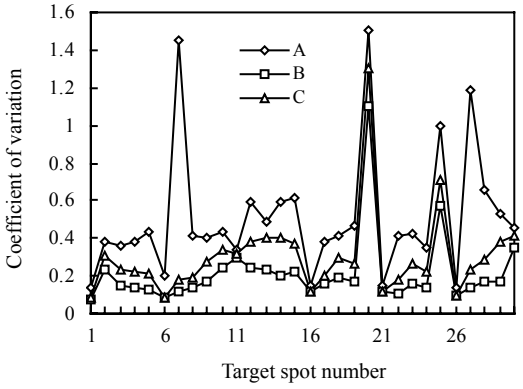



**Fig.11 Coefficients of variation of the target spots in protein microarray image**
A: original image (Fig.1); B: median-based filtered image (Fig.3); C: top-hat filtered image (Fig.6b)

## FURTHER DISCUSSION

The performance of the proposed enhancing method has been evaluated in another two kinds of protein microarray images, HBV (hepatitis B virus) and HLA (Human Leukocyte Antigen). The experimental results were very exciting. The SNR of the target spot improved significantly; and the

background became more uniform and the signal intensity was more consistent, just like the HCV microarray image we discussed before. The experimental results are shown in Fig.12.

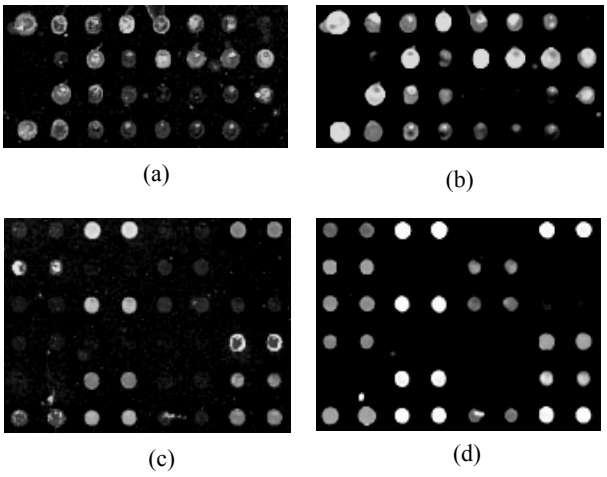



**Fig.12 Experimental results of HBV and HLA microarray image**
(a) original HBV image; (b) experimental result of (a); (c) original HLA image; (d) experimental result of (c)

For performance comparison, the standard mean filter, Gaussian filter, the proposed median based filter, the top-hat filter, and the proposed cascade filter were investigated. The window sizes of the filters are also discussed here. The results of SNR for the image in Fig.1 were shown in Fig.13.

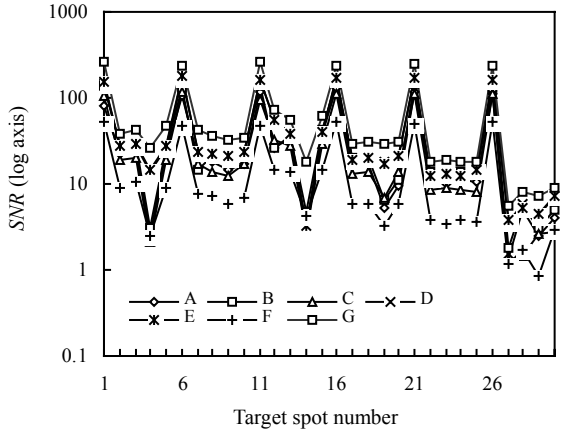



**Fig.13 SNR of each spot in Fig.1 after filtering by different filters**
A: 5×5 mean filter; B: 5×5 Gaussian filter; C: 3×3 median based filter; D: 5×5 median based filter; E: 7×7 median based filter; F: 21×21 top-hat filter; G: proposed cascade filter

The proposed 5×5 median based filter is better than the 3×3 one, and the 7×7 one is only a little better than the 5×5 one in SNR improvement, but it makes the edge of each target spot blurred and consumes more time. So we chose the 5×5 median based filter to reduce the impulsive noise in our experiments. After reducing the impulse noise, the contrast of the filtered image was enhanced by the top-hat filter. The SNR of each spot was improved more significantly than that of other filters. The background flatness and the consistency of signal intensity were also enhanced more evidently than those of other filters.

## CONCLUSION

This paper presents a novel algorithm of quality metric enhancement for protein microarray image. We consider the use of an adaptive median filter that employs the switching scheme based on local statistics characters to reduce noise levels in images and a top-hat filter to correct for trends in background values. The new cascade algorithm has been successfully applied to protein microarray images; we have evaluated the algorithm through three important factors that reflect the quality metric of protein microarray image: SNR of the spots, flatness of background and consistency of signal intensity. The experimental results revealed that this method significantly improved the SNR of target spots and the flatness of the background, but that it was a little inferior for improving the consistency of the signal intensity (when calculated by the coefficient of variation) compared to the one using a median-based filter.

## References

Axon Instruments, Inc., 2002. GenePix™ Pro 4.1 User's Guide and Tutorial, p.14-21.

Balagurunathan, Y., Dougherty, E.R., 2002. Simulation of cDNA microarrays via a parameterized random signal model. *Journal of Biomedical Optics*, **7**(3):507-523.

Brown, C.S., Goodwin, P.C., Sorger, P.K., 2001. Image metrics in statistical analysis of DNA microarray data. *Proceedings of the National Academy of Sciences*, **98**(16):8944-8949.

Chen, T., Wu, H.R., 2001. Adaptive impulse detection using center-weighted median filter. *IEEE Signal Processing Letters*, **8**(1):1-3.

Chen, Y., Kamat, V., Dougherty, E.R., Bittner, M.L., Meltzer, P.S.,Trent, J.M., 2002. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics*, **18**(9):1207-1215.

Glasbey, G.A., Ghazal, P., 2003. Combinatorial image analysis of DNA microarray features. *Bioinformatics*, **19**(2):194-203.

Jocelyn, H.N., Leodevico, L.I., 2002. Biomedical applications of protein chips. *Journal of Cellular and Molecular Medicine*, **6**(3):329-340.

Wang, Z., Zhang, D., 1999. Progressive switching median filter for the removal of impulse noise from highly corrupted images. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, **46**(1):78-80.

Wang, X., Ghosh, S., Guo, S., 2001. Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Research*, **29**(15):75-83.

Yang, Y.H., Buckley, M.J., Dudoit, S., Speed, T.P., 2002. Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, **11**(1):108-136.