A Comparison of x-lets in Denoising cDNA Microarray Images

Rouzbeh Shams[a] , *Student Member, IEEE,* Hossein Rabbani[b]*, *Senior Member, IEEE,* Saeed Gazor[c], *Senior Member, IEEE,*

[a] Electrical & Computer Engineering Department, Isfahan University of Technology, Isfahan, Iran

[b] Biomedical Engineering Department, Medical Image & Signal Processing Research Center, Isfahan University of Medical Science, Isfahan 81745319, Iran

[c] Electrical & Computer Engineering Department, Queen's University, Kingston, Canada

ABSTRACT

Microarray technology has become a power tool in the field of bioinformatics. It is used to measure gene expression levels and similar to any other image capturing processes is prone to noise. There are different kinds of noise, during preparation, hybridization and scanning in microarray images which usually are modeled by Gaussian noise. Since introduction of wavelets in 1970s, many more forms and extensions of this transform have been developed and used, such as stationary wavelet transform (SWT), complex wavelet transform (CWT), curvelet transform (CURV) and contourlet transform (CNT). By developing of more sparse transforms, it is important to have a perspective of how efficient the transforms are in different applications, such as microarray image analysis. In this paper, we compare the efficiency of common sparse transforms including ordinary discrete wavelet transform (DWT), SWT, CWT, CURV, CNT, Contourlet-SD decomposition, steerable pyramid (STP) and shearlet transform (SHR) for microarray image denoising. Therefore after converting microarray image into x-let transform, BayesShrink method, soft and hard thresholding are used to perform denoising of these images. Both local and general thresholds are calculated for each subband in order to evaluate the effect of incorporating intrascale dependency on top of sparsity property in statistical modeling of x-let's coefficients. Our simulation results show that CWT and SHR outperforms the others when using global thresholding and SWT is the preferred transform when using local thresholding. Although STP and SHR have better performance for some criteria like structural similarity (SSIM) index, but CWT is faster.

Keywords: microarray, denoising, x-let, transform, wavelet

1. INTRODUCTION

Microarray technology has become a powerful tool in the field of bioinformatics. It is used to simultaneously measure gene expression levels of thousands of genes and provide comprehensive genetic analysis of an organism or sample [1]. Due to the correlation of gene expression patterns and functions [2], this technology can provide valuable information for treatment of diseases such as cancer [3].

DNA arrays are two-dimensional substrates that different sequences of nucleotides are immobilized as spots on them in a certain pattern. Depending on their physical characteristics (such as diameters, density, etc.), these arrays are categorized into two categories: macro and micro arrays. Macroarrays are typically up to 22 cm long and carry approximately 1200 spots on each substrate. Microarrays however, are only a few centimeters in length and have the density of approximately 100,000 spots/cm^2 [1].

A typical Complementary DNA (cDNA) experiment consists of three main steps. In these experiments the goal is to determine the gene activity in two conditions: treatment and control [4]. The first step is preparing the samples which includes converting Messenger RNA to cDNA and labeling them with fluorescent dyes. Green Cy3 dye is used to mark the control and red Cy5 is used to mark the treatment sample. The next step is mixing these labeled samples and applying the mixture to the array for hybridization. The third step is to use scanners to obtain two scans of the microarray. One for each of the red and green channel. These images are then postporcessed, gene expression levels are calculated and analyzed [1,5,6].

These experiments, like any other, are prone to noise. Measurement of gene expression levels can be influenced by the noise introduced to the data during the preparation, hybridization and scanning phase. Additive or multiplicative Gaussian [7], poisson [8] and exponential [9] noise models have been used to describe the noise which affects microarray images and since non-additive Gaussian and other distributions here can be remodeled as Gaussian noise, in this study only Gaussian distribution is used to model the noise [4].

Reduction of noise, as in any other field, is of great importance in microarray technology. There have been studies regarding denoising microarray images, such as this work [10], in which the authors have used nonlinear filtering based on robust order statistics to remove noise from the image. Also in this study [11], a component based approach for reduction of noise in microarray images is proposed. In many works, the authors have used sparse transforms, e.g. Stationary Wavelet Transform (SWT) for removal of noise. For instance, in [12] denoising is done by smoothing the wavelet coefficients of the image. The authors of this study [13] used SWT, followed by thresholding to achieve the goal of enhancing microarray images by denoising. Similarly

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

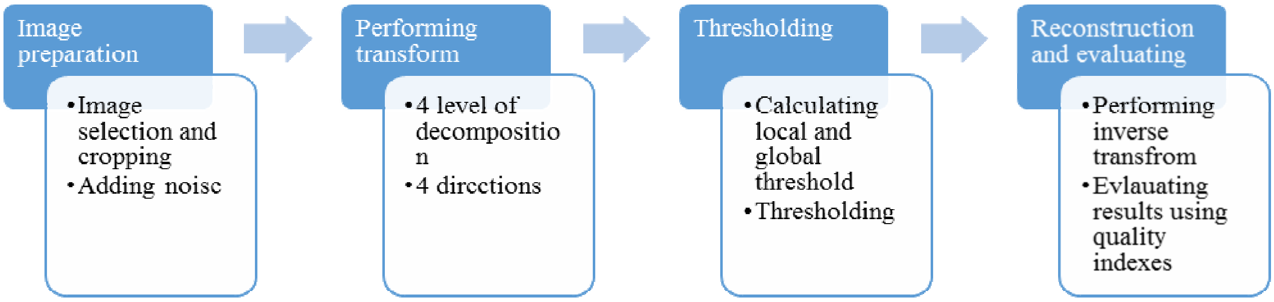

Figure 1 Denoising process

in [4], Complex Wavelet Transform (CWT) and estimation techniques are used in order to reduce the noise level of microarray images. As stated by authors of [4], denoising methods in transform domains have proven to be more efficient in noise reduction; that is if an appropriate transform is chosen.

Since introduction of wavelets in 1970s, many more forms of this transform have been developed and used, such as SWT, CWT, Curvelet transform, Contourlet transform, etc. By increasing application of sparse transforms in different fields including microarray image analysis, it is important to have a perspective of how efficient these transforms are in different applications such as denoising. Also because of the nature of microarray images (organized circular spots on a dark background), a transform that works better for denoising natural images, might not be as useful for denoising microarray images.

In this paper we will evaluate the performance of commonly used sparse transforms in denoising microarray images. We compare the results from Discrete Wavelet Transform (DWT), SWT, CWT, Curvelet transform, Contourlet transform and ControuletSD transform, steerable pyramid and shearlet transform using different quality indexes such as Peak Signal-to-Noise Ratio (PSNR), Structure Similarity (SSIM) and Edge Preservation Index (EPI). Generally, denoising is done by applying the transform on the image, thresholding, and applying inverse transform. Images used in this work are acquired from Stanford Microarray Data base [14] and UNC Microarray Database [15].

The rest of this paper is organized as follows. Section 2 explains the denoising process and quality measurement. In Section 3 the results and discussion is provided and in the final section conclusion and future works are given.

2. DENOISING PROCESS

The overall denoising process consists of four main steps: image preparation, performing transform, thresholding and performing inverse transform and evaluating results. These steps are common across all transformations and are explained below.

2.1 Image preparation

To prepare a benchmark for comparison of performance of different transforms, a dataset of synthetic noisy images was created. Images with low noise level were chosen for the process. Because there is no real data with zero noise level, images with low levels of noise, those that do not appear to be affected by noise, were chosen. These images were cropped to 512 by 512 pixel images and different levels of Gaussian noise were added to the images to evaluate the performance of the transforms at different noise levels.

2.2 Performing transform

After preparing images, each transform is performed on them. For each transform 4 levels of decomposition is used and in the case of directional transforms, such as CURV, CNT, etc., each level is decomposed into 4 directional subbands. For each transform, a toolbox was used for performing the transform. **Error! Reference source not found.** shows the list of the toolboxes used in this study.

2.3 Thresholding

After performing each transform, local and general thresholds are calculated for each subband using BayesShrink[16] method and soft thresholding is applied on the detail coefficients.

The threshold is calculated using the following equation

$$ts = \sigma_n^2 / \sigma_s \qquad (1)$$

where σ_s is standard deviation of the image without noise and σ_n^2 is the noise variance.

Table 1. The list of the toolboxes used

DWT	Matlab's DWT toolbox
SWT	Matlab's SWT toolbox
CWT	PolytechnicUni-DTCWT Toolbox, Cai, Li and Selesnick
Curvelet	Curvelab, Emmanuel Cand`es et al.
Contourlet	Contourlet Toolbox, Y. M. Lu and M. N. Do
Contourlet SD	ContourletSD Toolbox, Y. M. Lu and M. N. Do
Steerable Pyramid	MatlabPyrTools, E. P. Simoncelli
Shearlet	Shearlab, Wang-Q Lim et al

Table 2. Results of using hard thresholding.

Global Thresholding

	IPSNR = 34.15			IPSNR = 28.12			IPSNR =22.11			IPSNR = 17.25		
	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI
DWT	38.457	0.930	0.944	33.050	0.857	0.839	29.343	0.805	0.706	27.178	0.769	0.543
SWT	27.876	0.836	0.769	27.316	0.803	0.694	26.597	0.768	0.628	25.874	0.723	0.535
CWT	**38.627**	**0.932**	**0.946**	33.887	0.870	**0.867**	31.281	0.830	**0.811**	29.267	0.799	**0.696**
CNT	38.465	0.930	0.944	33.186	0.860	0.840	29.567	0.808	0.673	27.164	0.769	0.438
CNTSD	36.883	0.802	0.944	31.635	0.633	0.842	28.652	0.548	0.697	27.243	0.507	0.530
CURV	38.584	0.929	**0.946**	**33.943**	0.858	**0.867**	31.284	0.783	**0.811**	28.506	0.676	0.617
STY	38.584	0.933	0.945	33.787	**0.873**	0.861	31.113	0.833	0.786	**29.369**	**0.806**	0.659
SHR	38.577	**0.932**	**0.946**	33.843	0.870	0.862	**31.485**	**0.836**	0.801	29.097	0.801	0.643

Local Thresholding

	IPSNR = 34.15			IPSNR = 28.12			IPSNR =22.11			IPSNR = 17.25		
	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI
DWT	40.811	0.994	0.968	35.898	0.982	0.916	31.286	0.950	0.806	27.902	0.898	0.654
SWT	**44.361**	0.995	**0.985**	**39.828**	0.984	**0.963**	**35.562**	0.955	**0.916**	**32.287**	0.899	**0.842**
CWT	43.590	0.993	0.983	38.981	0.979	0.959	34.690	0.947	0.905	31.546	0.902	0.828
CNT	38.103	0.987	0.935	33.345	0.963	0.840	29.099	0.911	0.668	26.263	0.848	0.456
CNTSD	38.356	0.991	0.937	33.824	0.974	0.850	29.935	0.936	0.694	27.576	0.885	0.517
CURV	40.906	0.993	0.970	36.416	0.982	0.927	32.293	0.954	0.841	29.346	0.909	0.734
STY	42.391	**0.996**	0.976	37.906	**0.989**	0.940	33.716	**0.971**	0.863	30.666	0.940	0.752
SHR	41.196	0.995	0.967	37.233	0.988	0.927	33.597	**0.971**	0.854	30.964	**0.941**	0.769

The local variance for local threshold is found using a 3×3 sliding window on the noiseless image and the general threshold is calculated by taking the whole subband into account.

2.4 Reconstruction and evaluating

In the final stage, the images are reconstructed using the new coefficients.

Different image quality indexes including PSNR, SSIM, and EPI were used to determine the efficiency of each transform. PSNR is a common and well known quality index which can be calculated in the following manner:

$$\mathrm{PSNR} = 10.\, log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2}$$

which MSE is the Mean Squared Error and MAX_I is the maximum possible pixel value of the image.

SSIM is an index which shows structural similarities between two images and is found by the equation below.

$$\mathrm{SSIM} = \frac{(2\bar{s}\bar{\hat{s}} + 2.55)(2\sigma_{s\hat{s}} + 7.65)}{(\bar{s}^2 + \bar{\hat{s}}^2 + 2.55)({\sigma_s}^2 + {\sigma_{\hat{s}}}^2 + 7.65)} \tag{3}$$

where s, $\hat{s}$ are the original and denoised images, respectively. $\bar{s}$ Shows the mean of s, and $\sigma_{s\hat{s}}$ represents the covariance between s and $\hat{s}$.

EPI is an index used to measure edge preservation, that is, the more edges are preserved during the denoising process, the higher the index will be. EPI is measured using the following equation.

$$\mathrm{EPI} = \frac{\sum(\Delta s - \overline{\Delta s})(\Delta \hat{s} - \overline{\Delta \hat{s}})}{\sqrt{\sum(\Delta s - \overline{\Delta s})^2 \sum(\Delta \hat{s} - \overline{\Delta \hat{s}})^2}} \tag{4}$$

In the equation above, Δs is the highpass filtered s using the discrete Laplacian operator (a 3×3 pixel standard approximation) [17].

3. RESULTS

The results can be seen in Table 2 and 3, followed by examples of the cropped original image, noisy image and denoised image at the end of the paper (Figure 2). The maximum value of each column is highlighted.

As it can be seen from the results, while using hard thresholding and global threshold, CWT, SHR and STP seem to perform better than the others in terms of denoising. Using a local threshold the efficiency of SWT increases and makes it more successful in reducing noise.

In the case of soft thresholding using general threshold, CWT appears to be a better choice according to PSNR and EPI, but considering SSIM, STP has better performance than the rest. When local threshold is used however, SWT outperforms the rest.

4. CONCLUSION AND FUTURE WORK

In this paper, different transformations for denoising microarray images are evaluated/measured. According to our results, when using a global threshold, the PSNR and EPI of the output image for complex wavelet transform are the best. However, the STP and SHR have higher values for SSIM. As for the local threshold, the stationary wavelet transform was found to be the most effective.

For future works, the current comparison scheme can be extended to include the impact of the thresholding method on the results. It would be helpful if different methods finding the threshold were used. Moreover, because of the structure of these images, their histogram –either in pixel or transform domain- is predictable and by taking advantage of an efficient Bayesian estimator more effective noise reduction can be obtained.

REFERENCES

[1] U. Bilitewski, [Microchip Methods in Diagnostics], Humana Press, 1–14 (2009).

[2] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, "Cluster analysis and display of genome-wide expression patterns," Proc. Natl. Acad. Sci. USA 95 14863–14868 (1998).

[3] M. Esteller, P.G. Corn, S.B. Baylin, J.G. Herman, "A gene hypermethylation profile of human cancer," Cancer Res. 61 3225–3229 (2001).

[4] T. Howlader and Y. P. Chaubey, "Noise reduction of cDNA microarray images using complex wavelets.," IEEE Transactions on Image Processing, vol. 19, no. 8, pp. 1953–1967 (2010).

[5] E. Wit and J. McClure, [Statistics for Microarrays: Design, Analysis and Inference], 1st ed. Hoboken, NJ:Wiley, (2004).

[6] G. J. McLachlan, K. Do, and C. Ambroise, [Analyzing Microarray Gene Expression Data], 1st ed. Hoboken, NJ:Wiley, (2004).

[7] R. Lukac, K. N. Plataniotis, B. Smolka, and A. N. Venetsanopoulos, "A multichannel order-statistic technique for cDNA microarray image processing," IEEE Trans. Nanobiosci., vol. 3, no. 4, pp. 272–285, (2004).

[8] Y. Balagurunathan, E. R. Dougherty, Y. Chen, M. L. Bittner, and J. M. Trent, "Simulation of cDNA microarrays via a parameterized random signal model," J. Biomed. Opt., vol. 7, no. 3, pp. 507–523, (2002).

[9] S. W. Davies and D. A. Seale, "DNA microarray stochastic model," IEEE Trans. Nanobiosci., vol. 4, no. 3, pp. 248–254, (2005).

Table 3. Result of using soft thresholding.

Global Thresholding

	IPSNR = 34.15			IPSNR = 28.12			IPSNR =22.11			IPSNR = 17.25		
	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI
DWT	40.431	0.956	0.964	36.313	0.922	0.915	32.204	0.887	0.802	28.855	0.848	0.522
SWT	27.974	0.846	0.787	27.733	0.835	0.759	27.131	0.822	0.694	26.099	0.805	0.499
CWT	**41.009**	0.961	**0.969**	**37.355**	0.934	**0.938**	**33.365**	0.906	**0.858**	**29.740**	0.872	**0.633**
CNT	40.096	0.956	0.961	35.537	0.922	0.894	30.934	0.883	0.692	27.818	0.841	0.393
CNTSD	38.987	0.878	0.962	35.057	0.797	0.902	31.323	0.733	0.739	28.489	0.669	0.494
CURV	40.791	0.961	0.967	36.855	0.933	0.931	32.425	0.899	0.814	29.124	0.863	0.522
STY	40.895	**0.962**	0.968	37.094	**0.936**	0.933	32.922	**0.909**	0.832	29.692	**0.876**	0.610
SHR	40.607	0.959	0.965	36.769	0.933	0.924	32.541	0.905	0.817	29.266	0.872	0.566

Local Thresholding

	IPSNR = 34.15			IPSNR = 28.12			IPSNR =22.11			IPSNR = 17.25		
	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI	PSNR	SSIM	EPI
DWT	42.082	0.995	0.977	37.359	0.986	0.938	32.858	0.961	0.844	29.430	0.918	0.683
SWT	**45.411**	**0.998**	**0.987**	**40.623**	**0.995**	**0.968**	**35.793**	**0.986**	**0.919**	**31.960**	**0.969**	**0.824**
CWT	44.947	**0.998**	0.986	40.391	0.994	0.966	35.782	0.985	0.916	32.020	0.970	0.814
CNT	39.294	0.990	0.950	34.706	0.972	0.874	30.354	0.934	0.702	27.196	0.882	0.461
CNTSD	39.391	0.993	0.954	35.116	0.980	0.886	31.287	0.951	0.744	28.490	0.909	0.560
CURV	41.895	0.995	0.976	37.468	0.985	0.941	33.269	0.961	0.866	29.944	0.916	0.747
STY	43.469	0.997	0.982	39.052	0.991	0.954	34.851	0.975	0.889	31.613	0.947	0.784
SHR	42.307	0.996	0.975	38.115	0.989	0.941	34.014	0.971	0.875	30.698	0.938	0.782

[10] R. Lukac, K. N. Plataniotis, B. Smolka, and A. N. Venetsanopoulos, “A multichannel order-statistic technique for cDNA microarray image processing,” IEEE Trans. Nanobiosci., vol. 3, no. 4, pp. 272–285, (2004).

[11] A. A. Donald, Y. Zhang, and R. Parthe, “On denoising and com- pression of DNA microarray images,” Pattern Recognit., vol. 39, pp. 2478–2493, (2006).

[12] M. Mastriani and A. E. Giraldez, “Microarrays denoising via smoothing of coefficients in wavelet domain,” Int. J. Biomed. Sci., vol. 1, pp. 7–14, (2006).

[13] X. H. Wang, R. S. H. Istepanian, and Y. H. Song, “Microarray image enhancement by denoising using stationary wavelet transform,” IEEE Trans. Nanobiosci., vol. 2, no. 4, pp. 184–189, (2003).

[14] SMD.,http://smd.stanford.edu/index.shtml, 3 Jan. 2012.

[15] UNC.,< https://genome.unc.edu > (21 June 2012)

[16] Chang, S.G.; Bin Yu; Vetterli, M., "Adaptive wavelet thresholding for image denoising and compression," IEEE Transactions on Image Processing, vol.9, no.9, pp.1532,1546, (2000)

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans.Image Process., vol. 13, no. 4, pp. 600–612, (2004).

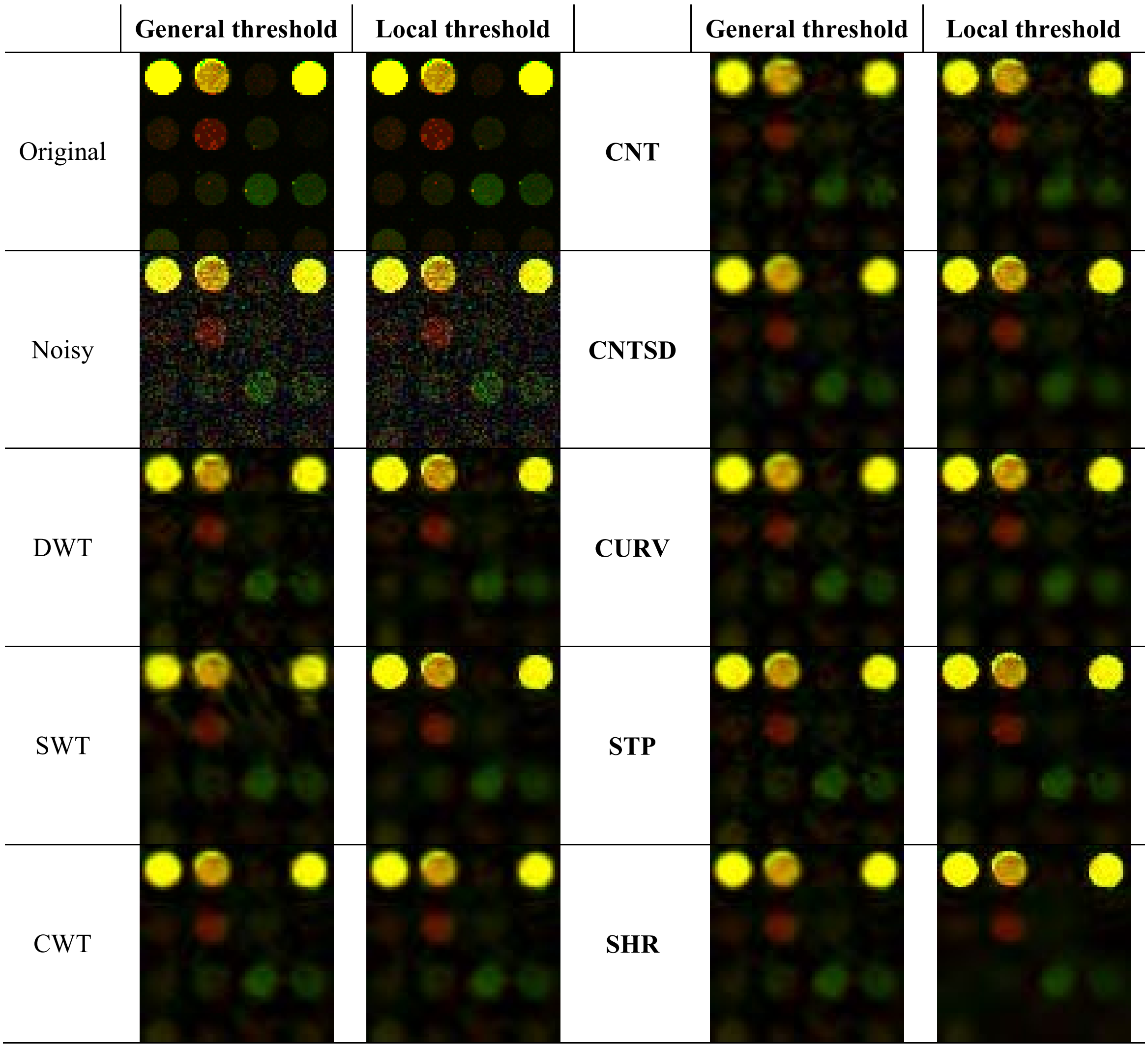

Figure 2. Example of original, noisy and denoised images, IPSNR = 17.25, Soft thresholding is applied.