

To cut or not to cut: can Graph Cutting Improve Microarray Gene Expression Reconstruction?

Karl Fraser, Zidong Wang, Yongmin Li, Paul Kellam* and Xiaohui Liu

Centre for Intelligent Data Analysis

School of Information Systems, Computing and Mathematics

Brunel University, Uxbridge, Middlesex, UB8 3PH, U.K.

*Department of Infection

University College London, London W1T 4JF, U.K.

Email: karl.fraser@brunel.ac.uk

Abstract—Microarrays produce high-resolution image data that are, unfortunately, permeated with a great deal of “noise” that must be removed for precision purposes. This paper presents a novel technique for such a removal process. On completion of this non-trivial task, a new surface (devoid of gene spots) is subtracted from the original to render more precise gene expressions. The Graph-Cutting technique as implemented has the benefit that only the most appropriate pixels are replaced. This means the influence of outliers and other artefacts are handled more appropriately (than in previous methods) as well as the variability of the final gene expressions being considerably reduced. Experiments are carried out to test the technique against commercial and previously researched reconstruction methods.

Index Terms—cDNA, Microarray, Background Reconstruction, Graph-Edge-Seam Cuts.

I. INTRODUCTION

Although microarray technology [1] was invented in the mid-90s’ the technology is still widely used in laboratories around the world today. The microarray “gene chip” contains probes for an organism’s entire transcriptome where differing cell lines render gene lists with appropriate activation levels. Gene lists can be analysed with application of various computational techniques, be they clustering [2] or modelling [3] for example such that the differential expressions can be translated into a clearer understanding of the underlying biological phenomena present. For a detailed explanation of the microarraying process readers may find references [4]–[7] of interest.

Addressing the issue of data quality effectively is however a major challenge, particularly when dealing with real-world data, as “cracks” will appear regardless of the design specifications etc. These cracks can take many forms, ranging from common artefacts such as hair, dust, and scratches on the slide, to technical errors like miscalculation of gene expression due to alignment issues or random variation in scanning laser intensity. Alongside these errors, there exist a host of biological related artefacts such as contamination of the complementary Deoxyribonucleic Acid (cDNA) solution or inconsistent hybridisation of multiple samples. The focus in the microarray field therefore is on analysing the gene expression ratios themselves [2], [8]–[12] as rendered from image sets. This means there is relatively little work directed

at improving the original images [13]–[16] such that final expressions are more realistic.

Noise in the images therefore will have a negative effect with respect to the correct identification and quantification of underlying genes. Therefore, in this paper we present an algorithm that attempts to remove the biological experiment (or gene spots) from the image. In the microarray field, it is accepted as part of the analysis methodology that the background domain (non-gene spot pixels) infringes on the gene’s valid measure and steps must be taken to remove these inconsistencies. In effect, this removal process is equivalent to background reconstruction and should therefore produce an image which resembles the “ideal” background more closely in experimental (gene spot) regions. Subtracting this new background image from the original would yield more accurate gene spot regions.

Gene expressions rendered by this reconstruction process are contrasted to those as produced by GenePix [17] (a commercial system commonly used by biologists to analyse images). The results are also compared with three reconstruction approaches (O’Neill *et al.* [14], Fraser *et al.* [15], [16]).

The paper is organised in the following manner. First, we formalise the problem area as it pertains to microarray image data and briefly explain the workings of contemporary approaches in Section II. Section III discusses the fundamental idea of our approach with the appropriate steps involved in the analysis highlighted. We then briefly describe the data used throughout the work and evaluate the tests carried out over both synthetic and real-world data in Section IV. Section V summarises our findings and gives some insight into future directions.

II. EXISTING TECHNIQUES

Microarray image analysis techniques require knowledge of a given gene’s approximate central pixel and the slide’s structural layout; therefore, all analysis techniques have similarities (regardless of their specific implementations). For example, a boundary is defined around the gene - thus marking the foreground region - with any outer pixels in a given radius taken to be local background. The background median is subtracted from the foreground and the result is summarised

as a \log_2 ratio. Bounding mechanisms include partitioning pixels via their histograms [8], [18], edge-based [19], [20], region growing [20], [21] and clustering [22], [23] functions, a detailed comparison of the more common approaches can be found in [13]. The underlying assumption throughout these mechanisms however is that there is little variation within the gene and background regions.

This is unfortunately not always the case as can be seen in the example regions of Fig. 1a, which depicts a typical test set slide (enhanced to show gene spot locations) with a total of 9216 gene regions on the surface held within an approximate area of $\sim 5000 \times 2000$ pixels. Note in addition that every image in the test set was created on a so called two-dye microarray system which means the DNA tagging agents are known as Cyanine 5 (Cy5) and Cyanine 3 (Cy3). The close-up sections provide good examples of the low-level signal produced in a typical image; problems such as partial or missing gene spots, shape inconsistencies, and background variation are clearly evident. Such issues are further highlighted in panels b and c where the scratch and background illuminations around the genes change significantly.

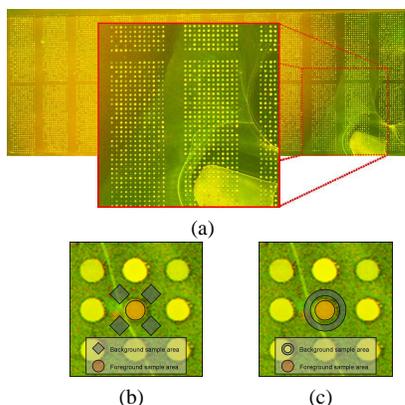


Fig. 1. Example Images: Typical test set Slide Illustrating Structure and Noise (a) with Sample Gene, Background Locations for GenePix Valleys (b) and ImaGene Circles (c)

A background identification process is required such that inherent variations between gene and background regions are handled more appropriately. Texture Synthesis represents one possible avenue for such reconstruction approaches as they deal with a similar problem. For example, Efros *et al.* [24] proposed a non-parametric reconstruction technique that is now well established. The underlying principal of the work was to grow an initial seed pixel (located within a region requiring rebuilding) via Markov Random Fields (MRF). Although this works well, the nature of the approach is such that speed is sacrificed for accuracy.

Bertalmio *et al.* [25] on the other hand took an approach inspired by the techniques as used by professional restorers of paintings; i.e. the principle of isotropic diffusion. Chan *et al.* [26] extended these works along with other related techniques and proposed an elastic curvature model approach that combined amongst others Bertalmio's transportation mechanism with the authors earlier Curvature Driven Diffusion work to produce accurate yet relatively slow reconstructions.

Oliveira *et al.* [27] tried to produce similar results to

Bertalmio, albeit much faster. Alas, microarray images contain tens of thousands of regions requiring such reconstructions and are therefore computationally expensive to examine with the aforementioned techniques.

Sun *et al.* [28] proposed an interactive method to inpainting with missing strong visual structures by propagating the structures according to user-specified curves which does improve on previous methods somewhat, but the interactivity clause would be inappropriate in this context.

Graph or seam-cutting on the other hand as used in the image-editing related field is quite prevalent. For example, Perez *et al.* [29] proposed a Poisson Image Editing technique to compute optimal boundaries between source and target images, while Agarwala *et al.* [30] created an interactive Digital Photomontage system that combined parts of a set of photographs into a composite picture. Kwatra *et al.* [31] proposed a system that attempted to smooth the edge between different target and source images. Other Graph related methods can be seen with [32]–[34] for example.

However, graph methods are predominantly designed for interactive usage (however minimal) while inpainting techniques are focused at producing aesthetic reconstructions rather than accurate ones as required in a medical context. To address the issue of removing objects, O'Neill *et al.* [14] attempted to harness ideas from the Efros *et al.* technique and improve background prediction results. Specifically, O'Neill *et al.* remove gene spots from the surface by searching known background regions and selecting pixels most similar to the reconstruction border. By making the new region most similar to given border intensities it is theorised that local background structures transition through the new region. However, the best such a process has accomplished in this regard is to maintain a semblance of valid intensities, while the original topological information is lost.

The next section describes an approach that attempts to address issues related to object removal by using a seam detection mechanism in an automatic and natural way.

III. A NEW TECHNIQUE

In this work, we have proposed Seam-Cut Image Reconstruction (*SCIR*), a novel technique that removes gene spots from a microarray image surface such that they are indistinguishable from the surrounding regions. Removal of these regions leads to more accurate gene spot intensities. Our previous work in this domain examined the effects of Recalibration (*HIR*) and Fourier Chaining (*CFIR*) (Fraser *et al.* [15], [16] respectively) techniques. Although *CFIR* dealt with shading and illumination issues more appropriately than *HIR*, *HIR* produced similar results significantly faster. However, both techniques can produce poorer reconstructions in regions dominated by strong artefacts (a saturated gene surrounded with similar level artefact for example). This work therefore attempts to improve on this issue; while at the same time generating exact pixel values (*CFIR* and *HIR* produce pixel estimates).

A. Description

The technique is designed to replace gene spot pixels with their most appropriate background neighbour. For example, a scratch on a photograph could be removed such that it is unidentifiable after reconstruction. In the context of this work, a scratch is equivalent to the gene spot region itself. Therefore, removal of this “scratch” should yield the underlying background region in the gene spot area. However, due to the nature of the microarraying process, gene spots can be rendered with different shapes and dimensions, individually and through the channel surfaces.

Therefore, we use a pre-defined window centred at a target gene (as determined by GenePix) to capture all pixels $p_{x,y}$ within a specified square distance from this centre. Note that (x,y) are the relative coordinates of the pixels in the window centred at pixel p . The Window size is calculated directly from an analysis of the underlying image along with resolution meta-data. The window can then be used to determine the appropriate *srcList* and *trgList* pixel lists (foreground and background) accordingly.

The gene spot pixels list can be defined via this windowed region as, $G^p = \Omega^w(g_{x,y})$, with Ω^w representing pixels falling into the windowed region and $(g_{x,y})$ meaning those pixels falling into the gene spot. The second list $B^p = \Omega^w(\bar{g}_{x,y})$ denotes those pixels within the same window that are not held in gene list G^p (and must therefore be representative of local background pixels).

The Seam-Cutting process then uses the *srcList* to determine those neighbouring pixels that have the strongest intensity through the surface. While *trgList* is used to determine the weakest neighbouring background intensities respectively. In the general sense, if we let image \mathbf{I} be a $n \times m$ surface, the vertical seam through the above lists could be defined as:

$$s = \{s_v\}_{x=1}^n = \{(v(x), x)\}_{x=1}^n, \forall x, |v(x) - v(x-1)| \leq m, \quad (1)$$

where x is the mapping $x:1,\dots,n;1,\dots,m$. The vertical seam is therefore an 8-way connected set of pixels in the image from top-to-bottom with one pixel per row. Initially, the image is parsed such that cumulative energy for all possible connected pixel sets is at a minimum for each x, y pairing through the surface.

In essence then, foreground pixels are replaced with their appropriate background equivalents. Such a replacement policy guarantees that the new foreground surface is not artificially biased to a particular intensity range. Indeed, if anything the new regions will consist of slightly lower intensity than perhaps is necessary meaning therefore a built-in buffer is also applied presently.

B. Example and Pseudo-Code

Initially, the SCIR process creates two distinct lists for a given gene spot location. The source list represents gene spot pixels as demarcated within the square window centred at the gene, while the target list consists of the remaining pixels in the window. Equation 1 is executed on the lists with the local background taken as the source region and the gene pixels the region to be reconstructed. Essentially the approach tries

to create a chain (or neighbouring set) of pixels through the region that have (in some sense) a minimal intensity. This can be thought of as a gradient function that searches for high-contrast (or edge) pixels within the gene spot region and low-contrast pixels within the local background region.

Fig. 2 presents a sample-reconstructed region from the Fig. 1a image as processed by the techniques. Note in particular how the SCIR surface looks sharper than that of O’Neill. This is due to the O’Neill surface being blurred such that resulting outliers etc are suppressed. The SCIR technique on the other hand generates absolute surfaces. A pseudo-code

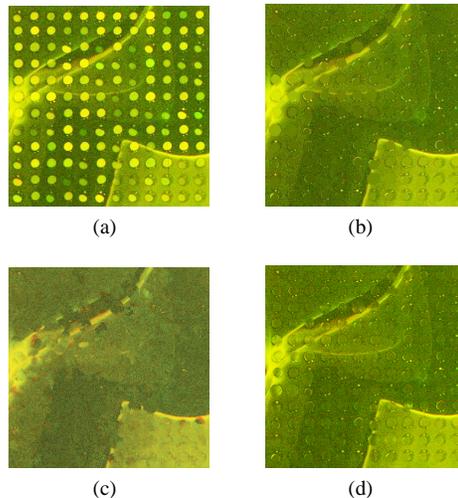


Fig. 2. Reconstruction Examples: Original Image (a), Reconstructed GenePix (b) O’Neill (c) and SCIR (d) Regions

implementation of the SCIR algorithm can be found in Table I. For clarity, the implementation is based on processing target window regions, which each contain a distinct set of pixels that are separated into gene spot and background sets.

IV. EXPERIMENTS AND RESULTS

This section details numerous experiments that were designed to empirically test the performance characteristics of the reconstruction methods. Median expression intensities are utilised in the comparisons as these values are in-fact the raw gene expressions (as used in post-analysis [2], [8]–[12] work for example). These values help provide clearer understanding of a gene spot’s repeat set and as such assist with clarification of the reconstruction quality itself.

A. Data set characteristics

The images used in this paper are derived from the human gen1 clone set [35] data. These experiments were designed to contrast the effects of two cancer-inhibiting drugs (PolyIC and LPS) over two different cell lines. One cell line represents the control (untreated) and the other the treatment (HeLa) line over a series of several time points. In total, there are 47 distinct slides with the corresponding GenePix results present. Each slide consists of 24 gene blocks with each block containing 32 columns and 12 rows of gene spots. The gene spots in the first row of each odd-numbered block

Input	
	<i>srcList</i> : List of gene spot region pixels
	<i>trgList</i> : List of sample region pixels
Output	
	<i>outList</i> : <i>srcList</i> pixels recalibrated into <i>trgList</i> range
Function seamCut(<i>srcList</i>,<i>trgList</i>):<i>outList</i>	
1.	For each gene
2.	geneRadius=radius of current gene spot
3.	While geneRadius not equal 0
4.	fgEnergy=calc max pixel surface from <i>srcList</i> members
5.	bgEnergy=calc min pixel surface from <i>trgList</i> members
6.	fgChain=Parse fgEnergy to determine max-neighbour pixel chain
7.	bgChain=Parse bgEnergy to determine min-neighbour pixel chain
8.	remove fgChain from fgEnergy
9.	remove bgChain from bgEnergy
10.	copy bgChain pixels into <i>srcList</i> locations
11.	geneRadius--1
12.	<i>outList</i> = <i>srcList</i>
13.	End While
14.	End For
End Function	

are known as the Lucidea ScoreCard [36], [37] and consist of a set of 32 pre-defined genes that can be used to test various experiment characteristics. The remaining 11 rows of the odd-numbered blocks contain the human genes themselves. The even-numbered blocks are repeats of their odd-numbered counterparts. This means that each slide has 24 repeats of the 32 ScoreCard genes and 4224 repeats of the human genes respectively. Note it is generally accepted that extreme pixel values should be ignored as these values could go beyond the scanning hardware’s capabilities.

B. Synthetic Data

The guiding principle of the technique is the feasibility that replacing gene spot pixels with pixels from neighbouring regions will result in a reconstructed area that is indistinguishable from the neighbouring region. Put another way, the gene spots should simply vanish from the surface which means that their new texture has to be very similar to the neighbouring region. Note that regions with strong and sharp intensity differences (an artefact edge for example) will be harder to “blend” successfully. In order to verify that the principle is at least valid, one would need to rebuild an obscured known region and compare before and after surfaces for accuracy. However, as the gene spot sits above the optimal background surface it is not possible to determine optimal rebuild pixels. In order to validate rebuild feasibility therefore, we use the Synthetic Gene Spot (SGS) creation process as outlined in Fraser *et al.* [16].

The first experiment therefore is focused at answering “how well the SCIR process removes synthetic gene spots from the image”? Sixty-Four (64) realistic SGS’s were placed into existing background regions of the Fig. 1a images Cy5 and Cy3 surfaces. These synthetic gene’s were then reconstructed with the before and after surfaces compared for similarity. Note that as the artefact region itself could be considered gene spot similar, our reconstruction processes also attempt to build the region such that the artefact pixels are removed. This process yields a ball-park-figure for the potential distillation errors generated by the various background reconstruction techniques. Such potentials as rendered from test imagery can be seen in Fig. 3a, while Fig. 3b highlights a close up sample region of the aforementioned SGS’s.

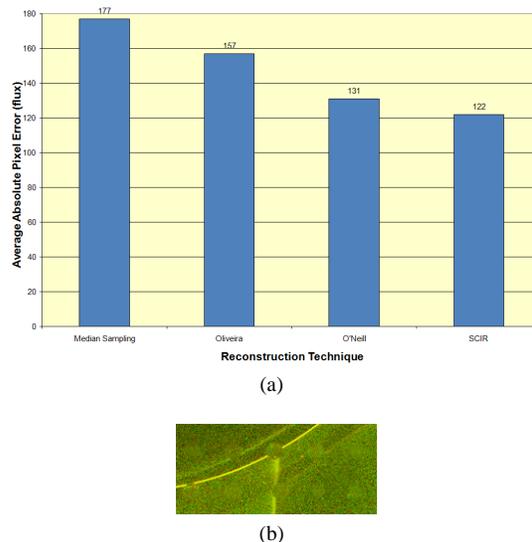


Fig. 3. Synthetic Gene Spots: Average Absolute Pixel Error (a) and close up of a Fig. 1a region with ten synthetic spots (b)

The graph presents the potential intensity flux error (PIFE) for the reconstruction techniques. On average, the GenePix advocated median sampling approach yields a PIFE of 177 per pixel per SGS region while the other techniques yield decreasing values (our process value of 122 represents a ~30% reduction over GenePix). Such a finding reiterates that downstream analysis when based on GenePix (specifically the BackGround Correction (BGC) stage) estimates directly; produce more erroneous gene expressions than perhaps appreciated.

The panel b surface highlights a sample of the SGS region with a large artefact running through two (2) gene spot regions. Note (as stated above) we can see that the strong artefact edge has been successfully replaced with appropriate background substitutions. Note however that such strong edges can cause greater challenges within real data as shall be seen.

C. Real Data

With our confidence in the reconstruction techniques abilities enhanced by the synthetic results, the next stage is to understand how such reconstructions fare with real data. In particular, “how badly do strong artefact edges interfere with a reconstruction event”?

Experiment two only uses the ScoreCard control genes for all blocks across the test images. Recall, the composition of the test imagery is such that we have more technical repeats of the control genes than the human ones. Also, the control genes are completely independent of the biological experiment which means ideally they should fluoresce in exactly the same way across the images regardless of environmental conditions (in principle).

The Fig. 4 plot presents the tracking of the standard deviations (STD) for the 32 ScoreCard genes over the 24 repeat locations. Note however that due to the way in which O'Neill calculates a given gene spots region, their STD's are somewhat lower than expected. However, the plot still imparts general characteristics for the given reconstruction techniques.

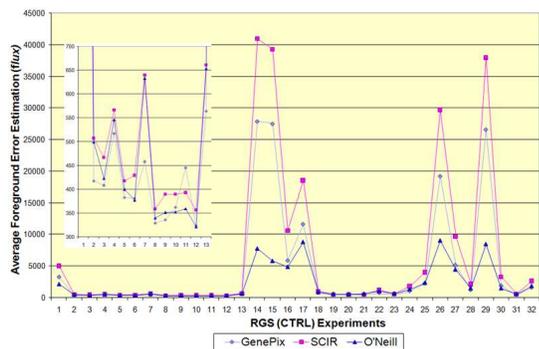


Fig. 4. Real Gene Spots: Overview of ScoreCard gene Standard Deviations

If we disregard saturated gene spots for a moment and examine the close up section of the plot we see the profile residuals follow each other fairly well. This means the processes do reduce STDs at least in a partial sense.

Critically then, this leads to a need to understand the reconstruction techniques performance characteristics more closely. Specifically the relationships between expression measurements for all ScoreCard genes in the same slide (Fig. 5a) and across all slides in the test set (Fig. 5b) are compared. Note that it is expected that some intensity differences will appear as the experimental time point's increase as required through the biological processes.

These plots show the bound absolute foreground median values for the multiple image channels for the documented techniques. From Fig. 5a it can be seen that SCIR and O'Neill performed in a similar vein with very little difference amongst them. However, the saturated gene spots - 15 in this case has caused a blip in the profile plot for SCIR. Recall, that by the very nature of a saturated gene spot, the surface is close to a constant value and obviously artificially high. But in this instance the gene in question also has a strong artefact intercepting it. During reconstruction, the constant type value of the gene is not a major challenge to rectify; more problematic is how to deal with the strong intercepting artefacts appropriately. Note that the replacement pixel sets as derived during reconstruction actually do a fair job overall. For this image, the saturated gene spots did not affect the outcome of the final quantification stage greatly.

Whereas the Fig. 5a plot represents a specific image surface, which does not render a given reconstruction techniques

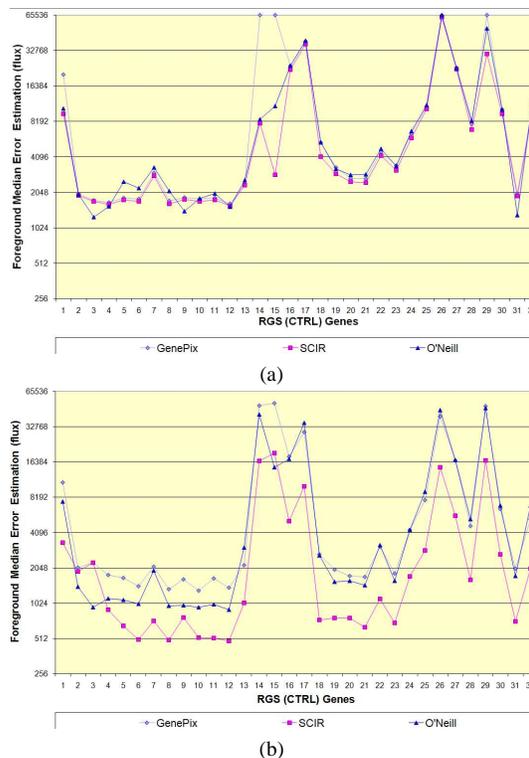


Fig. 5. Real Gene Spots: Absolute Medians for 32 genes over Fig. 1a and Test set (b) Regions

abilities to deal with a range of image modalities. The panel b plot therefore shows the same information across the entire 47-slide test set. This should allow us to see how exactly a reduction manifests itself onto the final gene metrics. Clearly the SCIR process has reduced the technical repeats to a greater extent than perceivable from the sample image alone. The respective profile values for the test set are 10374, 3742 and 9213 flux respectively.

Clearly, reconstruction of gene spot's does have a positive effect on the final expression results but, not so obvious, are the ramifications that the reconstruction has over the test set. Fig. 6a therefore is a comparison chart showing explicitly the improvement (or not) of a particular reconstruction technique against the original GenePix expressions.

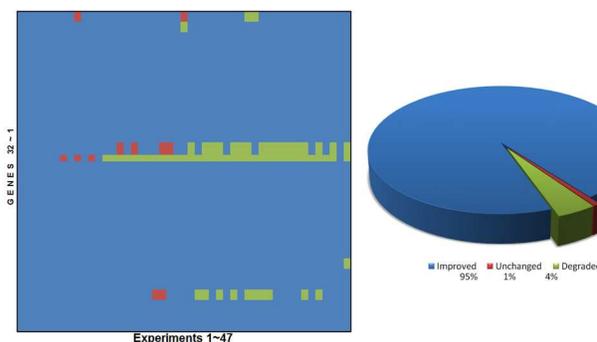


Fig. 6. Final Results Comparison: Matrix for test set showing difference in repeat expression fluctuations; the GenePix, SCIR and O'Neill techniques are assigned the colours blue(darkest), red and green(lightest) (~10% difference) respectively

The general gene region 16~17 banding and 30 (partial banding) as seen in Fig. 6 are associated with aforementioned saturated (or near background) gene intensities as created by the Axon scanner hardware and are suggestive of more work needed. The non-banded genes on the other hand are indicative of the individual reconstruction techniques being able to account more appropriately for gene intensity replacement.

V. CONCLUSIONS

The paper looked at the effects of applying current and new texture synthesis inspired reconstruction techniques to real-world microarray image data. In particular, we proposed a novel approach to reconstructing a gene's underlying background by attempting to focus on problematic pixels only. In our previous work we have looked at trying to harness an image's global knowledge more closely while at the same time being constrained by local conditions for example. Although that process was shown to be highly effective, there were still weaknesses of the system. In this work we attempted to address some of these weaknesses more closely and as the results show the new method makes a significant improvement in gene repeat variance reduction.

Although in future we would like to be able to compare our reconstruction processes with other mainstream analysis methods like Spot [13], ScanAlyze [38], ImaGene [39] and QuantArray [18] for instance, it is not only difficult to acquire appropriate results, but internal result workings of the methods are also needed. In addition, it is quite probable that a hybrid reconstruction system (able to classify to some extent a gene region) will be of great benefit to this analysis task. Such a hybrid system would use what is deemed to be the most appropriate reconstruction technique for a given gene.

ACKNOWLEDGMENT

This work is in part supported by EPSRC grant (EP/C524586/1) and an International Joint Project sponsored by the Royal Society of the U.K. and the National Natural Science Foundation of China.

REFERENCES

- [1] D. Shalon and R. Davies, "Quantitative monitoring of gene expression patterns with a complementary dna microarray," *Science*, vol. 250, no. 5235, pp. 467–470, 1995.
- [2] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," in *Proceedings of the National Academy of Sciences, USA*, December 1998, pp. 14 863–14 868.
- [3] P. Kellam, X. Liu, N. Martin, C. A. Orengo, S. Swift, and A. Tucker, "A framework for modelling virus gene expression data," *Journal of Intelligent Data Analysis*, vol. 6, no. 3, pp. 265–280, 2002.
- [4] B. Coe, "You want ketchup with your dna chips? an overview of expression microarrays," *BioTeach Online Journal*, vol. 1, no. 1, pp. 89–94, 2003.
- [5] G. M. Duyk, "Sharper tools and simpler methods," *Nature Genetics*, vol. 32, pp. 474–479, 2002.
- [6] E. F. Petricoin III, J. L. Hackett, L. J. Lesko, R. K. Puri, S. I. Gutman, K. Chumakov, J. Woodcock, D. W. Feigal Jr., C. K. Zoon, and D. F. Sistiare, "Medical applications of microarray technologies: a regulatory science perspective," *Nature Genetics*, vol. 32, pp. 474–479, 2002.
- [7] X. Liu and P. Kellam, "Mining gene expression data," in *Bioinformatics: Genes, Proteins and Computers*, 1st ed., C. A. Orengo, D. T. Jones, and J. M. Thornton., Eds. Oxford: BIOS Scientific Publishers, 2003, vol. 0, no. 15, pp. 229–244.

- [8] Y. Chen, E. R. Dougherty, and M. L. Bittner, "Ratio-based decisions and the quantitative analysis of cdna microarray images," *Journal of Biomedical Optics*, vol. 2, pp. 364–374, 1997.
- [9] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown, "Genomic expression program in the response of yeast cells to environmental changes," *Molecular biology of the cell*, vol. 11, pp. 4241–4257, 2000.
- [10] J. Quackenbush, "Computational analysis of microarray analysis," *Nature Reviews Genetics*, vol. 2, no. 6, pp. 418–427, June 2001.
- [11] B. M. Kepler, L. Crosby, and T. K. Morgan, "Normalization and analysis of dna microarray data by self-consistency and local regression," *Genome Biology*, vol. 3, no. 7, p. research0037.1, 2002.
- [12] J. Quackenbush, "Microarray data normalization and transformation," *Nature Genetics*, vol. 32, pp. 490–495, 2002.
- [13] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed, "Comparison of methods for image analysis on cdna microarray data," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 108–136, 2002.
- [14] P. O'Neill, G. D. Magoulas, and X. Liu, "Improved processing of microarray data using image reconstruction techniques," *IEEE Transactions on Nanobiotechnology*, vol. 2, no. 4, 2003.
- [15] K. Fraser, Z. Wang, Y. Li, P. Kellam, and X. Liu, "Improving microarray expressions with recalibration," in *American Institute of Physics. Utrecht, The Netherlands: AIP*, October 4th 5th 2007, pp. 3–16.
- [16] —, "Noise filtering and microarray reconstruction via chained fouriers," *Advances in Intelligent Data Analysis VII*, vol. 7, no. 1, pp. 308–319, 2007.
- [17] Anonymous, *GenePix Pro Array Analysis Software*, Axon Instruments Inc.
- [18] —, *QuantArray Analysis Software*, GSI Lumonics.
- [19] W. A. Perkins, "Area segmentation of images using edge points," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 2, no. 1, pp. 8–15, 1980.
- [20] N. Ahuja, A. Rosenfeld, and R. M. Haralick, "Neighbour gray levels as features in pixel classification," *Pattern Recognition*, vol. 12, pp. 251–260, 1980.
- [21] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641–647, 1994.
- [22] J. McQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability*. Berkeley, CA: L. Le Cams and S. Neyman (Eds), 1967, pp. 281–297.
- [23] L. Kaufman and J. P. Rousseeuw, "Finding groups in data: An introduction to cluster analysis," in *Clustering Large Applications (Program CLARA)*, L. Kaufman and J. P. Rousseeuw., Eds. New York: John Wiley and Sons, 1990, no. 3, pp. 126–163.
- [24] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *IEEE International Conference on Computer Vision*, 1999, pp. 1033–1038.
- [25] M. Bertalmio, A. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *IEEE Computer Vision and Pattern Recognition*, December 2001.
- [26] T. Chan, S. Kang, and J. Shen, "Euler's elastica and curvature based inpaintings," *Journal of Applied Mathematics*, vol. 63, no. 2, pp. 564–592, 2002.
- [27] M. M. Oliveira, B. Bowen, R. McKenna, and Y. S. Chang, "Fast digital image inpainting," in *Proceedings of the Visualization, Imaging and Image Processing*, Marbella, Spain, September 2001, pp. 261–266.
- [28] J. Sun, L. Yuan, J. Jia, and H. Shum, "Image completion with structure propagation," 2005, pp. 861–868.
- [29] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 313–318, 2003.
- [30] A. Agarwala, M. Dontcheva, M. Agrawala, s. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 294–302, 2004.
- [31] V. Kwatra, A. Schdl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: Image and video synthesis using," San Diego, California, July 2003, pp. 277–286.
- [32] W. A. Barrett and A. S. Cheney, "Object-based image editing," San Antonio, Texas, July 2002, pp. 777–784.
- [33] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut - interactive foreground extraction using iterated graph cuts," Los Angeles, California, August 2004, pp. 309–314.
- [34] F. Nielsen and R. Nock, "Clickremoval: interactive pinpoint image object removal," Hilton, Singapore, November 2005, pp. 315–318.
- [35] H. G. M. Project., "Human gene1 clone set array."

- [36] H. Samartzidou, L. Turner, T. Houts, J. Frome, M. Worley, and H. Albertsen, "Lucidea microarray scorecard: An integrated analysis tool for microarray experiments," *Life Science News*, vol. 7, no. 13, pp. 1–10, 2001.
- [37] "Stability studies of dyes in microarray applications, Tech. Rep. 1, 2003.
- [38] Anonymous, *ScanAlyze Array Analysis Software*, Eisen Labs.
- [39] —, *ImaGene Array Analysis Software*, BioDiscovery Inc.