Microarray Image Processing: Current Status and Future Directions

COMPLEMENTARY DNA microarray imaging is considered an important tool and powerful technology for large-scale gene sequence and gene expression analysis. DNA microarrays represent a technological intersection between biology and computers that enables gene expression analysis in human on a genome-wide scale. In general, thousands of gene-specific probes are arrayed on a small matrix, such as a glass slide or microchip, and this matrix is probed with labeled nucleic acid synthesized from a tissue type, development stage, or other condition of interest. The expression profiles of thousands of genes under that condition can, thus, be assayed simultaneously [14]. In particular, this technique is now considered a powerful tool for studying the genetic basis of complex diseases.

The origin of this imaging technique was evolved from Southern's technique in the 1970s [1] and sparked in the last decade. The microarray experiment produces thousand of genes on a single slide, and all the genes can be globally viewed at the same time. This provides a systematic and comprehensive way to survey the DNA and RNA variations [2], which could become a standard tool for both molecular biology research and genomic clinical diagnosis, such as cancer diagnosis [3] and [4] type 1 [6] and type 2 [6] diabetes diagnosis.

The DNA sequences are spotted on the slides. After further image processing, gene expression data can be obtained for further analysis, such as gene clustering or identification. Cluster analysis can identify the genes that are regulated in a similar manner under a number of experimented conditions. DNA arrays can also be used to characterize the cellular differences between different tissue types, such as between normal cells and cancer cells, or between cancers with different responses to treatment, or between control cells and cells treated with a particular drug.

The data drawn from a DNA chip are mainly fluorescent images organized into a matrix of spots, whose intensity is propotional to specific, site-dependent, DNA hybridization. One of the major advantages of this technique is the parallelism of the process. Hence, with just one experiment it is possible to collect a large number of the relevant data necessary for the genomic analysis. However, one of the major challenges of this approach is the image processing phase. The purpose of this phase is to extract each spotted DNA sequence as well as to obtain background estimates and quality measures. The accuracy of this phase has substantial impact on the accuracy and effectiveness on the subsequent gene expression and identification analysis. Hence, image processing techniques are becoming fundamental elements in this process in order to speed up and enhance the accuracy of extracting the information contained in these images.

The processing of the microarray images can be separated into three sequential steps, namely, gridding, segmentation, and intensity extraction, while segmentation is supposed to be the most important step. Over the last few years, a few commercial and freeware softwares have been developed, such as ScanAlyze [7], GenePix [8], and QuantArray [9]. These softwares use either simple segmentation methods or advanced methods such as *watershed* [10], [11] and *seeded region growing* [12], [13] to recognize and extract the gene intensity. But the problems still exist for processing poor quality images, such as noisy and weak spot images. Advanced technologies are also required to obtain new high-quality gene expression methodologies [14].

Some of the major areas of investigation and future challenges that can be tailored for future research directions in this emerging area are the development of new and advanced intelligent image processing techniques for both eliminating the noise sources inherent in the DNA microarray process and also the development of tailor-made image processing methodologies for speeding up the real-time diagnosis and implementation procedures of the next generation of system-on-a-chip devices. These can also target the provision of compatible networking structures for these systems to link these systems, for example, with existing powerful grid computers for future chronic disease management. These can ultimately provide a new generation of diagnostic systems that can help to unlock the unknown patterns of complex diseases and their molecular phenotypes and allow rapid and responsive treatment mechanisms for these diseases.

In this special issue, we have selected six papers to address some of these challenges in enhanced imaging processing techniques and the new methodologies addressed above. In the first paper, O'Neill *et al.* discuss an image reconstruction technique to eliminate the noise sources in microarray images. The paper proposes an processing technique that can take a mask of the artifact and then recreate the image of the noise. Subsequently, this noise image is subtracted from the original image to leave the remaining signal. The genes can then be calculated in a normal fashion by measuring the intensity of the remaining pixels of each gene. This technique can reduce the noise dramatically. The first paper by X. H. Wang *et al.* discusses an image enhancing technique by denoising using stationary wavelet transform. This technique can overall eliminate the random noise to improve the quality of gene expression data. For poor-quality image arrays, X. H. Wang *et al.* present a new spot recognition method based on wavelet modulus maxima. It is well known that wavelet transform has the ability to recognize the discontinuities and edges of the images, and is very sensitive to the weak signal. The methodology presented in this paper allows the enhanced recognition of invisible spots and can identify the correct size. The technique of third dye array visualization is
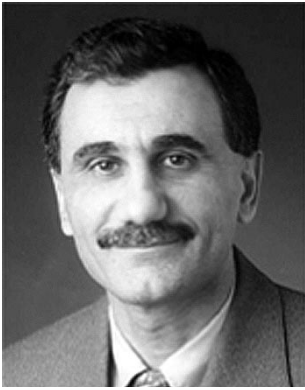
presented by Xujing Wang *et al.* to improve the quality of microarray imaging data. The technique presented in this paper enables direct prehybridization of the printed array. One of the significant advantages of this technique is that the spot intensities on the fluorescein image are reflective of the amount of solid surface-bound probe. The fluorescein image offers much greater homogeneity in spot signal intensity. Furthermore, the third dye prehybridization image also provides a criterion for judging which coordinates should contain signal information after hybridization, as only those that have spotted materials on them should be considered a candidate for data acquisition. The paper by Katzer *et al.* describes image processing method for automatic spotted microarray image analysis. The paper describes a Markov random field based approach to high-level grid segmentation. This robust approach is applied to several images to validate the methodology used. The last paper, by Bozinov *et al.*, presents a Java-based software technique to provide an enhanced access methodology for remote users of mircoarray imaging technologies.

In conclusion, I hope you will find this issue interesting and useful. I also hope that this topic will be a major research area that can span multidisciplinary domains and provide a synergy between future topics in bioinformation engineering and e-med systems.

ROBERT S. H. ISTEPANIAN
School of Computing and
    Information Systems
Mobile Information and
    Network Technologies Centre
Kingston University
London, U.K.

REFERENCES

[1] E. M. Southern, "Detection of specific sequences among DNA fragments separated by gel electrophoresis," *J. Mol. Biol.*, vol. 98, pp. 503–517, 1975.

[2] E. S. Lander, "Array of hope," *Nature Genetics Suppl.*, vol. 21, pp. 3–4, Jan. 1999.

[3] A. Alizadeh, M. Eisen, P. O. Brown, and L. M. Staudt, "Probing lymphocyte biology by genomic-scale gene expression analysis," *J. Clin. Immunol.*, vol. 18, pp. 373–379, 1998.

[4] C. M. Perou, S. S. Jeffrey, M. Van de Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu, J. C. F. Lee, D. Lashkari, D. Shalon, P. O. Brown, and D. Botstein, "Distinctive gene expression patterns in human mammary epithelial cells and breast cancers," in *Proc. Nat. Acad. Sci.*, vol. 96, 1999, pp. 9212–9217.

[5] L. M. Scearce, J. E. Brestelli, S. K. McWeeney, C. S. Lee, J. P. Mazzarelli, F. Deborah, A. S. Pizarro, C. J. Stoechert Jr., S. P. Sandra, M. A. Permutt, J. Brown, A. Douglas, and K. H. Kasestner, "Functional genomics of the endocrine pancreas: The pancreas clone set and PancChip, new resources for diabetes research," *Diabetes*, vol. 51, pp. 1997–2004, 2002.

[6] S. T. Nadler, J. P. Stoehr, K. L. Schueler, G. Tanimoto, B. S. Yandel, and A. D. Attie, "The expression of adipogenic genes is decreased in obesity and diabetes mellitus," *Proc. Nat. Acad. Sci.*, vol. 97, pp. 11 371–11 376, 2000.

[7] M. B. Eisen. ScanAlyze. [Online]. Available: http://rana.stanford.edu/software

[8] "GenePix 4000A User's Guide," Axon Instruments Inc., Union City, CA, 1999.

[9] *QuantArray Analysis Software, Operator's Manual*, 1999.

[10] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," in *Mathematical Morphology in Image Processing*. New York: Marcel Dekker, 1993, vol. 34, Optical Engineering, pp. 433–481.

[11] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 583–598, June 1991.

[12] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 641–647, June 1994.

[13] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed, "Comparison of methods for image analysis on cDNA microarray data," Statistics Dept., Univ. Cal., Berkeley, Tech. Rep. No. 584, 2000.

[14] A. Baxevanis and B. F. Ouellette, *Bioinformatics, A Practical Guide to Analysis of Genes and Proteins*, 2nd ed. New York: Wiley, 2001.

Robert S. H. Istepanian (M'91–SM'97) received the Ph.D. degree in electronic and electrical engineering from Loughborough University, Loughborough, U.K., in 1994.

From 1984 to 1998, he worked in different overseas industrial and academic positions. In 1988, he was a Visiting Research Fellow in the Department of Electronic and Electrical Engineering, Loughborough University. From 1994 to 1995, he was with the same department as a Postdoctoral Research Fellow. From 1996 to 1999, he was a Senior Lecturer at the University of Portsmouth, Portsmouth, U.K. From 1999 to 2000, he was an Associate Professor at Ryerson University, Toronto, ON, Canada, and also an Adjunct Professor with the University of Western Ontario, London, ON, Canada. He was also the Head of the Mobile Information Engineering and E-Med Systems research group in the Department of Electronic and Computer Engineering, Brunel University, London, U.K. He is currently Professor of Data Communications in the School of Computing and Information Systems and Director of the Mobile Information and Network Technologies Research Centre (MINT), Kingston University, London, U.K. He is Recipient, Investigator, and Coinvestigator of research grants from the Engineering and Physical Science Research Council, British Council, and Royal Society on intelligent computational systems applied to telesurgery systems. He is also a lead partner on several EU-funded projects on e-health. Other collaborative and funded projects from industry (Vodafone, Orange, and Motorola UK) in the areas of m-health systems and microarray imaging technologies are also ongoing. He has published more than 120 refereed journal and conference papers, mostly in the areas of of wireless telemedicine and biomedical signal processing, and has edited three books in the areas of m-health, biomedical communications, and signal processing. His current research lies in the cognate areas of wireless communications and e-health and advanced coding theory for genomics and microarray imaging technologies.

Dr. Istepanian currently serves on the advisory board of IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, and was one of the Founding Special Area Editors on mobile telemedicine. He was also the Guest Editor of the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE Special Issue on Mobile Telemedicine Systems (Sept. 2000), and is Guest Editor of the forthcoming Special Issue on m-Health to be published in late 2003. He was the Co-Chair of the UK/RI chapter of the IEEE Engineering in Medicine and Biology Society (EMBS). He has also served on numerous technical committees, and has chaired and was an invited speaker for several national and international IEEE conferences, including the Telemed conferences of the Royal Society of Medicine, the Annual International Conference of the EMBS (EMBS'98 and '99), and the 2000 World Medical Congress. He was also the Technical Co-Chair of the 2003 IEEE EMBS Conference on Information Technology and Applications in Biomedicine (ITAB) and the Fifth International Workshop on Enterprise Networking and Computing in HealthCare (HealthCom'03). His biography is included in *Who's Who in Science and Engineering* (New Providence, NJ: Marquis, 2003).