In most cases, the direct downsampling method outperforms JPEG, without downsampling. This is because only a quarter of the original data is needed to compress in the downsampling compression scenario. The IDID-based image compression scheme significantly outperforms the former two. This can be attributed to the property that IDID is able to preserve more information in the downsampled images.

Fig. 7 gives the visual comparisons among different compression methods at 0.20 bpp for Elaine (512×512). It is shown that JPEG results in the worst visual quality due to the existence of severe blocking artifacts. In the reconstructed image by Direct + NLEDI (direct downsampling and NLEDI interpolation), blocking artifacts disappeared; however, there is much more noise. This is because a lot of detail information is lost during direct downsampling, and it cannot be recovered by NLEDI. On the contrary, the reconstructed images by IDID_Bilinear + Bilinear (IDID_Bilinear downsampling and Bilinear interpolation) and IDID_NLEDI + NLEDI (IDID_NLEDI downsampling and NLEDI interpolation) exhibit better visual quality and higher PSNR.

V. Conclusion

An IDID algorithm has been proposed in this paper. Different from other downsampling algorithms, the proposed IDID hinges the interpolation to the downsampling process. For each input image, the IDID is able to obtain an optimal downsampled image from which a high-visual-quality image with the same resolution as the input image is generated. We have also proposed a content-dependent IDID algorithm for the interpolation methods with varying interpolation coefficients. Experimental results demonstrate the viability and efficiency of the proposed IDID.

References

[1] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down scaling for better transform compression," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1132–1144, Sep. 2003.

[2] Y. Tsaig, M. Elad, P. Milanfar, and G. H. Golub, "Variable projection for near-optimal filtering in low bit-rate block coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 154–160, Jan. 2005.

[3] W. Lin and D. Li, "Adaptive downsampling to improve image compression at low bit rates," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2513–2521, Sep. 2006.

[4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[5] X. Wu, X. Zhang, and X. Wang, "Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 552–561, Mar. 2009.

[6] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[7] S. H. Jung, S. K. Mitra, and D. Mukherjee, "Subband DCT: Definition, analysis, and applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 273–286, Jun. 1996.

[8] D. Mukherjee and S. K. Mitra, "Image resizing in the compressed domain using subband DCT," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 620–627, Jul. 2002.

[9] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[10] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 6, pp. 508–517, Dec. 1978.

[11] Q. Wang and R. Ward, "A new orientation-adaptive interpolation method," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 889–900, Apr. 2007.

[12] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[13] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[14] X. Zhang, S. Ma, Y. Zhang, L. Zhang, and W. Gao, "Nonlocal edge-directed interpolation," in *Proc. IEEE Pacific-Rim Conf. Multimedia*, Dec. 2009, pp. 1197–1207.

[15] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, Jun. 2005, pp. 60–65.

[16] T. Frajka and K. Zeger, "Downsampling dependent upsampling of images," *Signal Process. Image Commun.*, vol. 19, no. 3, pp. 257–265, Mar. 2004.

[17] J. Goutsias and H. J. A. M. Heijmans, "Nonlinear multiresolution signal decomposition schemes—Part I: Morphological pyramids," *IEEE Trans. Image Process.*, vol. 9, no. 11, pp. 1862–1876, Nov. 2000.

[18] H. J. A. M. Heijmans and J. Goutsias, "Nonlinear multiresolution signal decomposition schemes—Part II: Morphological wavelets," *IEEE Trans. Image Process.*, vol. 9, no. 11, pp. 1897–1913, Nov. 2000.

[19] Y. Zhang, J. Zhang, R. Xiong, D. Zhao, and S. Ma, "Low bit-rate image coding via interpolation oriented adaptive down-sampling," in *Proc. SPIE, Visual Commun. Image Process.*, Jul. 2010, p. 77 441C.

[20] *Spatial Scalability Filters*, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Jul. 2005.

Cellular Neural Networks, the Navier–Stokes Equation, and Microarray Image Reconstruction

Bachar Zineddin, Zidong Wang, and Xiaohui Liu

***Abstract*—Although the last decade has witnessed a great deal of improvements achieved for the microarray technology, many major developments in all the main stages of this technology, including image processing, are still needed. Some hardware implementations of microarray image processing have been proposed in the literature and proved to be promising alternatives to the currently available software systems. However, the main drawback of those proposed approaches is the unsuitable addressing of the quantification of the gene spot in a realistic way without any assumption about the image surface. Our aim in this paper is to present a new image-reconstruction algorithm using the cellular neural network that solves the Navier–Stokes equation. This algorithm offers a robust method for estimating the background signal within the gene-spot region. The MATCNN toolbox for Matlab is used to test the proposed method. Quantitative comparisons are carried out, i.e., in terms of objective criteria, between our approach and some other available methods. It is shown that the proposed algorithm gives highly accurate and realistic measurements in a fully automated manner within a remarkably efficient time.**

***Index Terms*—cDNA microarray reconstruction, cellular neural networks (CNN), isotropic diffusion, Navier–Stokes equations (NSEs), partial differential equations (PDEs).**

I. Introduction

DNA microarray is a remarkably successful high-throughput technology for functional genomics [23]. Microarrays allow researchers to collect quantitative data about the expression level of many thousands of genes in a single experiment. Therefore, it offers a deep understanding of gene interaction and regulation. However, the microarray

Manuscript received October 13, 2010; revised April 07, 2011; accepted May 24, 2011. Date of publication June 09, 2011; date of current version October 19, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rafael Molina.

The authors are with the Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, U.K. (e-mail: bachar.zineddin@gmail.com; Zidong.Wang@brunel.ac.uk; Xiaohui.Liu@brunel.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2159231

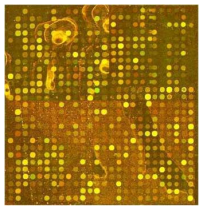

Fig. 1. Part of a microarray image gives a good example of the variations in the background signal. The target is to reconstruct the areas of the spots, i.e., the bright circle signals, assuming that there is no information in these areas ("0"s).

is still far from perfection, and it needs improvements in all the main stages of the microarray process, including image analysis.

Although microarray technology has been engineered to a fine tolerance, there exists high signal variability through the surface of the microarray image (see Fig. 1). Due to this imperfection in the microarray production, many types of noise, i.e., biological or systematic [11], contaminate the resulting image, and it becomes a challenging issue to address the microarray data quality effectively. Although the image-analysis stage significantly affects the identification of differentially expressed genes [14], the researchers tend to focus their efforts to tackle the challenges of gene expression ratios as rendered from the microarray images [7]. So far, a few studies have been dedicated to the analysis of the image itself [11].

Generally, analyzing the microarray image complies with the following steps [24]: 1) the filtering stage; 2) the gridding stage; 3) the segmentation stage; and 4) the final stage, which is the quantification of the underlying genes. The currently available software packages for image analysis (e.g., see [1] and [10]) are largely dependent on manual, semiautomated, or fully automated methods, which consume a lot of processing time. For instance, the ScanAlyze [10] software requires the operator to take as many as 14 steps [3], and many of these steps have to be repeated several times. Therefore, in order not to lose the advantages of the parallelism of the microarray technology, a hardware implementation of microarray image analysis is an auspicious alternative to these tools. Furthermore, the hardware utilization should process the image to obtain highly accurate and realistic data in a fully automated manner within a remarkably efficient time.

In order to overcome the aforementioned bottleneck for microarray image processing, Samavi *et al.* [22] proposed a hardware architecture to analyze the microarray image. In particular, Arena *et al.* [2] used cellular neural networks (CNNs) and analog-and-logical (analogic)-signal-dedicated CPU called CNN universal machine. Note that the CNN [9] framework provides a flexible approach to describe spatiotemporal dynamics in the discrete space. In particular, it allows for an efficient very-large-scale-integration implementation of analog array-computing structures. Such devices possess huge processing power that can be employed to solve numerically expensive problems. The CNN representation of a partial differential equation (PDE) is a spatially discrete dynamical system, which is qualitatively equivalent to the original spatially continuous system. Both systems operate in continuous time, and the values of state variables, interactions, and parameters are all continuous. In [2], the proposed algorithm facilitates the parallel nature of the CNN to achieve the required objectives. Unfortunately, there have been two limitations with the methods developed in [2] and [22]. First, the operator has to define, in advance, a specific set of threshold values in order to address the intensity analysis. Consequently, the analysis depends on a particular hypothesis about the underlying question of the experiment. Second, in these methods, it is assumed implicitly that the background noise and other artifacts are absent in the output of the segmentation stage, which is not necessarily the case. To this end, it is concluded that the image-analysis stage should have a specific background determination process that can analyze the inherent variation between the gene and the background signals within any proposed hardware framework dedicated to microarray image analysis, and this constitutes the motivation of our current investigation.

In this paper, a new methodology for cDNA microarray image reconstruction is proposed. The idea is to use a practical CNN approximation to solve the Navier–Stokes equation (NSE), which describes the fluid velocity in the incompressible fluid, i.e., to obtain an exemplary approximation of the background in the gene-spot region. The theoretical basis of this approach can be found in [5], where the remarkable similarity has been highlighted between the steam function and image intensity. It has also been suggested in [5] that the NSE solution is applicable for the inpainting (reconstruction) purpose. A CNN is an analogic processor array that allows the application of a local strategy, i.e., with less computational complexity, to meet the task requirements. It is important to note that using local information leads to a robust and reliable algorithm in some applications such as microarray image reconstruction as we will see later. Due to its architecture, the 2-D CNN array is used widely to solve image-processing and pattern-recognition problems. Furthermore, the parallelism of this structure allows one to perform the most computationally expensive image-analysis tasks in a faster way than a classical CPU-based computer. For this paper, subtracting the reconstructed background from the original spot signal should give rise to more accurate quantification of genes' signals.

The main contribution of this paper lies in two aspects. First, a new microarray image-reconstruction algorithm is proposed by using the CNN that solves the NSE, and such an algorithm is proven to be robust for estimating the background signal within the gene-spot region. Although the relation between Bertalmio's reconstruction method and the NSE has been established and a CNN algorithm has been investigated to achieve the NSE approximation in the incompressible fluid studies, a full CNN algorithm for image reconstruction based on the NSE has not been developed yet and remains challenging. In addition, the special characteristics of the microarray images make this task far from a straightforward application. Second, the CNN templates (complete set of templates) are developed with specific steps to achieve the microarray image reconstruction. This paper is organized in the following manner. First, we formalize the problem area as it pertains to microarray image data and briefly explain some available approaches in Section II. Section III discusses the basic idea of our proposed algorithm with appropriate steps involved in the analysis highlighted. We then briefly describe the data used throughout this paper and evaluate the algorithm over real-world data in Section IV. Section V summarizes our findings and renders some observations into possible future directions.

II. Existing Image-Reconstruction Techniques

Regardless of the microarray image-analysis methodologies that have been followed, all of them deal with the same basic principles. For example, the features' location and the classification of the pixels into foreground and background signals have to be carried out. Then, the median of the gene spot and the median of the background pixels are taken to be foreground and background representatives, respectively. Typically, in the existing literature, by assuming that there is a little variation within the gene and background regions, the background median is subtracted from the foreground, and the result is summarized as a $\log 2$ ratio. Unfortunately, this is not necessarily true. For instance, many problems can exist, such as missing or partial gene spots, shape inconsistencies, and background variation such as the scratch and the variation of the background illuminations around the presented genes.

In the past years, there has been a growing need for establishing a more specific background determination process that can account for the inherent variation between the gene and background regions. One of the first techniques applied specifically to reconstruct microarray images is from O'Neill *et al.* [17]. In particular, a gene area is replaced by

selected pixels that are most similar to the known border from a known background region. Most recently, Zineddin *et al.* [25] proposed a CNN algorithm to predict the pixels' values in the reconstructed area based on the boundary information. The underlying assumption is that the similarity with the given border intensities guarantees the transition of the local background structures through the new region. However, the algorithm in [25] has a parameter that should be specified based on the image surface. In addition, the algorithm utilizes one-layer boundary information to fill the reconstructed area, and it would be interesting to improve the inpainting method by employing more information about the layer, which is a series of pixels forming a closed line.

III. Novel CDNA Microarray Image-Reconstruction Technique

A. Description

In this paper, we propose to use a CNN approximation to the NSE for image reconstruction (CNN-NSIR), which is a novel technique that removes gene-spot regions from a microarray image surface. The removal of these regions leads to a more accurate background estimation, which can then be used to yield yet more realistic genes' signal. Techniques such as that in [17] in the spatial domain exclusively and essentially utilize the gene border pixels as reference values to produce appropriate pixel mappings. Although this approach works well, such a kind of brute-force methods is typically expensive with respect to the execution time. However, if we can utilize the locally spatial information integrated with a hardware implementation, we are able to overcome this limitation. Suppose we have subset $\Omega \in D$ where we would like to modify the gray level of I based on the information of I from the surrounding region $D \setminus \Omega$, where Ω is the reconstructed region. The modified region I^*, i.e., the solution, will have equal values as I in $D \setminus \Omega$. The process of finding the appropriate I^* is the reconstruction problem.

The approach proposed in [4] attempts to mimic techniques as used by skilled artists to perform inpainting manually. It works on the principle of a PDE isotropic diffusion model. Using a mask to specify the area to be inpainted, the algorithm fills in these areas by propagating the information of the border region along a level line (isophotes). Isophotes are level lines of equal gray levels. Mathematically, the direction of the isophotes can be interpreted as $\nabla^{\perp} I$, where ∇ is gradient (∂_x, ∂_y) and $\nabla^{\perp} = (-\partial_y, \partial_x)$ means the direction of the smallest change. Next, the smoothness could be interpreted as ΔI, where Δ is the usual Laplace operator $(\partial_x^2 + \partial_y^2)$. In general, ΔI will extract an edge and noise in an image. Therefore, in order to mimic the idea of artistic inpainting, we should propagate ΔI in the direction of $\nabla^{\perp} I$ from the boundary of the reconstructed area $\partial\Omega$. Consequently, the solution criterion for the inpainting problem I^* satisfies $(\nabla^{\perp} I^* \cdot \nabla \Delta I^* = 0)$, and it is equal to I on $\delta\Omega$, which is the boundary of Ω. However, microarray images contain thousands of regions requiring such reconstructions and therefore are computationally expensive to examine with the highlighted technique. In an attempt to handle such a time restriction, Oliveira *et al.* [16] aimed to produce similar results to [4] albeit quicker, although the approach may lead to loss of some information in the translation.

In [5], an inpainting approach has been introduced based on the ideas from classical fluid dynamics to propagate isophote lines continuously from the border into the reconstructed region. The underlying assumption is to think of the image intensity as a stream function for a 2-D incompressible flow. The Laplacian of the image intensity plays the role of the vorticity of the fluid, i.e., it is propagated into the inpainted area by a vector field defined by the stream function. The method is based directly on the NSEs for fluid dynamics, which have the immediate advantage of well-developed theoretical and numerical results. The basic equation for the incompressible Newtonian flow is as follows:

$$\frac{\partial \mathsf{v}}{\partial t} + \mathsf{v}\nabla\mathsf{v} = -\nabla p + \mu\nabla^2\mathsf{v} \tag{1}$$

where v is the velocity vector, p is the pressure, and μ is the viscosity. For 2-D flows, we introduce the stream function Ψ, where $\nabla^{\perp}\Psi = \mathsf{v}$.

In image-processing terms, we have the counterpart to the vorticity-stream function, i.e.,

$$\omega_t + \mathsf{v}\nabla\omega = \mu\nabla \cdot (g(|\nabla\omega|)\nabla\omega) \tag{2}$$

where $\Delta I = \omega$ is the vorticity, $\nabla^{\perp} I = \mathsf{v}$ is the direction of the isophotes, and $g(.)$ accounts for anisotropic diffusion (or edge-preserving diffusion).

B. Designing the Templates

Again, let Ω be a small area to be reconstructed (inpainted) and let $\partial\Omega$ be its boundary. The small size of the gene spot Ω allows the ability to use isotropic diffusion in order to propagate information from one or two layers of pixels from the boundary of the gene spot $\partial\Omega$ into Ω. Therefore, approximating inpainting procedure has been achieved.

As a basic framework, let us consider a 2-D $M \times N$ CNN array in which the cell dynamics is described by the following nonlinear ordinary differential equation with linear and nonlinear terms:

$$\begin{aligned} C\frac{d}{dt}x_{ij}(t) &= -R^{-1}x_{ij}(t) + \sum_{kl\in N_r} A_{i,j;k,l}y_{kl}(t) \\ &\quad + \sum_{kl\in N_r} B_{i,j;k,l}u_{kl} + z_{ij} \\ y_{ij}(t) &= 0.5\left(|x_{ij}(t)+1| - |x_{ij}(t)-1|\right) \end{aligned} \tag{3}$$

where $|x_{ij}(0)| \le 1, |u_{ij}| \le 1, |z_{ij}| \le z_{\max}$ $(1 \le i \le M, 1 \le j \le N)$, and x_{ij}, u_{ij}, and y_{ij} are the state, input, and output voltages of the specific CNN cell, respectively. The state and the output vary in time, and the input is static (time independent) (for more details about CNN see [8], [9], and [19] and the references therein).

In all the following template, h is the uniform grid size, and R is the value of the state resistor in a CNN cell. In addition, provided that the transient remains bounded (i.e. the cells do not saturate), it is assumed that a CNN array is stable when it starts from a specified initial condition.

The first derivative I_x can be mapped directly onto the CNN array, resulting in the following simple template $(\text{DER}x : X(0) = \text{ORIGINAL IMAGE},\ \text{BC} = \text{ZF})$:

$$\text{DER}x_A = [0\ 0\ 0; -1/2h\ 0\ 1/2h; 0\ 0\ 0],\ DERx_B = 1/R. \tag{4}$$

Similarly, the first derivative $-I_y$ can be mapped directly onto the CNN array, resulting in the following simple template $(\text{DER}y_- : X(0) = \text{ORIGINAL IMAGE},\ \text{BC} = \text{ZF})$:

$$\text{DER}y_{-A} = [0\ 1/2h\ 0; 0\ 0\ 0; 0\ -1/2h\ 0],\ \text{DER}y_{-B} = 1/R. \tag{5}$$

The linear isotropic diffusion equation can be mapped directly onto the CNN array, resulting in the following simple template [9] $(\text{DIFFUS} : X(0) = \text{ORIGINAL IMAGE}, \text{BC} = \text{ZF})$:

$$\text{DIFFUS}_A = [0\ 1/h^2\ 0; 1/h^2\ -4/h^2 + 1/R\ 1/h^2; 0\ 1/h^2\ 0]. \tag{6}$$

There are a considerable number of methods for numerical integration. One of the best known techniques is the Newton–Côtes method that is based on a polynomial interpolation on equally spaced points. This method can be transformed into integration rules using polynomials of any order, giving an error that decreases faster and faster with the number of points being used as higher order polynomials are chosen [15]. In the light of the work of Luchini [15] and by applying the closed Newton–Côtes formula (Simpson's rule),

we can compute numerically the integration over the x-axis (or y-axis) based on the spatial information and the following template (INTx : $X(0)$ = ORIGINAL IMAGE, BC = ZF):

$$\text{INT}x_A = [0\ 0\ 0; 0.4h\ 0\ 0.4h; 0\ 0\ 0],\ \text{INT}x_B = 2h + 1/R. \quad (7)$$

Finally, a CNN approximation of the NSE (2) can be created using a two-layer CNN (see the solution of NSE for incompressible fluids in [13]). With the developed rational in [13], we can evaluate v in (1) using the CNN template (8) for the u component (those for the v component can be generated analogously) (NSE : $X(0)$ = ORIGINAL IMAGE, BC = ZF) as follows:

$$\begin{aligned}\text{NSE}A_{uu} &= [0, \mu/h^2, 0; \mu/h^2, -4\mu/h^2 + 1/R, \mu/h^2; \\ &\qquad 0, \mu/h^2, 0] \\ \text{NSE}\hat{A}_{uu} &= [0, 0, 0; -1/2h, 0, 1/2h; 0, 0, 0] \\ \text{NSE}\hat{A}_{uv} &= [0, -1/2h, 0; 0, 0, 0; 0, 1/2h, 0]. \end{aligned} \quad (8)$$

Remark 1: All the derived templates have been tuned to be stable in the grayscale area (i.e., the linear area in the output function, see [8, Chapter 6]). Therefore, after a specific transient time elapses, the output of any specific operation would be either the state value or the output value.

C. Pseudocode of CNN-NSIR

As discussed previously, a common drawback with the existing microarray image-processing methods is that they cannot address properly the quantification of the gene spot in a realistic way without any assumption about the image surface. In order to overcome this drawback, a CNN algorithm, i.e., CNN-NSIR, is proposed in this paper with the pseudocode given in algorithm 1.

Algorithm 1 CNN-NSIR Algorithm

Require Ω {Image: specify gene-spot region pixels}
Require I {Image: the image to be reconstructed}
Ensure I_o {Image: the output image}
$\{\Omega \in \{0,1\}\}$ $\{I \in [-1,1]\}$

1: $\Omega \leftarrow$ CNNDilation$(\Omega, 2)$ {Add two pixel layer to Ω}
2: $I \leftarrow$ CNNMask$(I, \hat{\Omega})$ {$\hat{\Omega}$ is The Complementary of Ω}
3: $I \leftarrow$ CNNDiffusion(I)
4: Set α {α: # of NSE evaluations}
5: Set β {β: transient time of NSE evaluation}
6: $\grave{\Omega} \leftarrow$ CNNDilation$(\Omega, 1)$
7: $u \leftarrow$ CNNget_u$(I, \partial\grave{\Omega}, DERy_)$
8: $v \leftarrow$ CNNget_v$(I, \partial\grave{\Omega}, DERx)$
9: **for** $i = 1$ to α **do**
10: $[U, V] \leftarrow$ CNNnse$(u, v, \grave{\Omega}, \beta)$ {Propagate $\partial\grave{\Omega}$ into $\grave{\Omega}$}
11: $I_1 \leftarrow$ -CNNIntegration$(U, INTy)$
12: $I_2 \leftarrow$ CNNIntegration$(V, INTx)$
13: $I \leftarrow I_2 - I_1$
14: $\grave{\Omega} \leftarrow$ threshold(I) {returns 1 where cells' values equal 0}
15: $\grave{\Omega} \leftarrow$ CNNDilation$(\grave{\Omega}, 1)$
16: $u \leftarrow$ CNNget_u$(I, \partial\grave{\Omega}, DERy_)$
17: $v \leftarrow$ CNNget_v$(I, \partial\grave{\Omega}, DERx)$
18: **end for**
19: $I_o \leftarrow I$

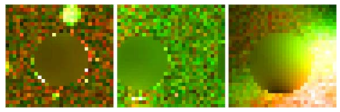

Fig. 2. Samples of three different background signals. The reconstructed images show an acceptable background trend estimation. The third example represents the effect of the noise artifacts on the spot's area.

Essentially, the proposed algorithm takes mask Ω and the input image I. Note that, in this algorithm, we deal with each channel separately; thus, we should consider I as either the Cy_5 or Cy_3 image. The mask is an image that marks the spot regions with pixels of value "1." The first step in the algorithm is to add a layer of two-pixel thick. This layer guarantees the elimination of the effect of the direct spot boundary pixels. In the segmentation result, the boundary pixels usually contain overlapping information from the spot and the background signal. Therefore, it is not a good representative of the background signal in the local area. Fig. 2 presents a sample-reconstructed region.

IV. Discussion

A. Notes About the Algorithms

Due to the high signal variability that exists across the microarray surface, when working directly with the raw microarray information, propagating the information of the border region $\partial\Omega$ along a level line (isophotes) would be impotent. Therefore, rather than using the raw image, it is suggested that producing a smoothed version of the image data would not only be advantageous but also more effective in terms of the overall goal. In our algorithm, the DIFFUS template has been used as a smoothing operator. Although this isotropic diffusion template causes blurring effects, it can achieve the required refining result by calculating the region (local) average intensities.

Anisotropic diffusion would be a better alternative for the diffusion operator not only as a smoothing operator (see line 3 in algorithm 1) but also in evaluating the NSE as well. However, the anisotropic diffusion models require the noise-level estimate K that determines the magnitude of the edges to be conserved during the smoothing process. K could be set according to the *a priori* knowledge about the noise statistics or could be estimated from the absolute gradient histogram [6], [18]. However, in a locally connected parallel processing architecture, it would be very difficult to calculate these values. Thus, other approaches should be sought to achieve this target. Rekeczky *et al.* [19] proposed a possible method to estimate the noise level roughly as the minimum of the maximal local variations at the nodes of the coarse-grid model.

To evaluate the outcome of the proposed approach, the results have been compared with the output image by two other algorithms that are dedicated specifically for reconstructing (inpainting) images. These two algorithms are Bertamio's method [4] and the "f-Inpaint" method [25]. Although the isotropic diffusion operator has been applied as a first step in Bertalmio's algorithm (and anisotropic diffusion thereafter), the microarray characteristics do cause a very long settling time. The "f-Inpaint" method uses the information of single-bit thickness layer to reconstruct the spot's area. Although the "f-Inpaint" method is remarkably faster than CNN-NSIR, it causes more spots to be considered as a bad region and therefore omitted in later analyses.

B. Data Set Characteristics

The images used in this paper are derived from the human gen1 clone set 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 presented. 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 are known as the Lucidea ScoreCard [21] and consist of a set of 32 predefined genes that can be used to test various experiment characteristics.

C. Evaluation

In order to quantify the performance capabilities of our technique, a quality measure is required to allow the judgment of how the estimated background affects the quantification of the gene spot. For this purpose, a systematic objective method, which is based on the descriptive statistical interclass correlation coefficient (ICC) measures [12], is used to compare the results produced by different techniques. The rational is justified as follows. The set of 32 predefined genes is used in the comparison process. Using these controls, we base our analysis on the following assumptions: 1) The better the reconstruction is, the higher the correlation within the same control should be (minimum σ_e^2); 2) the better the reconstruction is, the lower the correlation between the genes within the array should be (maximum σ_g^2); and 3) the better the reconstruction is, the higher the ICC value should be.

In order to compare the proposed algorithm with a commonly used median background estimation, we estimate the reliability of each method for each experiment using the components of the variance model $Y_{ij} = g_i + e_{ij}$, where Y_{ij} is the log-expression ratio for the i spot and the j replicate. The error variance component σ_e^2, which is associated with e_{ij}, represents the reproducibility of the method, i.e., how much the spots signals within several replicates differ from the real gene's expression. The variance component σ_g^2, which is associated with g_i, represents the true spot-to-spot (gene-to-gene) variability, i.e., the variance of all spot signals from the real mean of all genes' expressions. Then, the intraclass correlation coefficient (ICC) represents the reliability of the method. The ICC is used as a measure of reproducibility over a measure such as the error variance or its square root σ_e alone because it guards against algorithms that produce ratio estimates that are all shrunk to a central value.

The variance component, which is the error within genes and between replicates, is estimated by $\hat{\sigma}_e^2 = \sum_{i=1}^{n_g}\sum_{j=1}^{n_a}(Y_{ij} - \bar{Y}_{i\cdot})^2/[n_g(n_a - 1)]$, where n_a is the number of replicate arrays, n_g is the number of genes, and $\bar{Y}_i = \sum_{j=1}^{n_a} Y_{ij}/n_a$ is the mean of every individual gene's log-expression ratios over the replicate arrays. The between-gene variance component is estimated by $\hat{\sigma}_g^2 = \sum_{i=1}^{n_g}(\bar{Y}_i - \bar{Y}_{\cdots})^2/(n_g - 1) - \hat{\sigma}_e^2/n_a$, where $\bar{Y}_{\cdots} = \sum_{i=1}^{n_g}\sum_{j=1}^{n_a} Y_{ij}/(n_g n_a)$ is the mean of all genes in all arrays. The estimated ICC is ICC $= \hat{\sigma}_g^2/(\hat{\sigma}_g^2 + \hat{\sigma}_e^2)$.

Fig. 3 presents the estimated variance components and the ICC for the data set images and on average. The reliabilities of all methods are high, with the CNN-NSIR method appearing on average to be more reliable than the other methods. Note that even the within-spot variability $\hat{\sigma}_e^2$ (the noise) is notably smaller for the CNN-NSIR method, although the between-spot variability (the signal) is bigger for the f-Inpaint method. Bertalmio's method with a smoothing step at its beginning works very well for most images. When it comes to the high-throughput microarray images, Bertalmio's method cannot be applied directly as the processing becomes very time consuming. In fact, our proposed algorithm improves the applicability of Bertalmio' method in terms of reducing the processing time. It is shown that our algorithm could facilitate potentially the hardware implementation of Bertalmio's method in order to reconstruct the microarray image.

It should be pointed out that the blind algorithm (i.e., an operator-independent algorithm) limits our ability to discriminate between bad spots from the good ones. However, this will give a better insight about

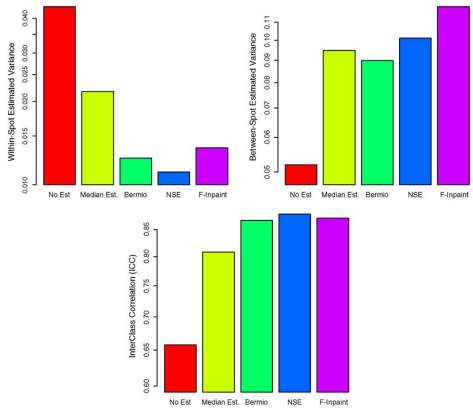

Fig. 3. "Y-axes": (a) Within-spot estimated variance. (b) Between-spot estimated variance. (c) Average ICC over the data set. ("X-axes:" The methods are "No Background Correction," "Median Background Estimation," "Bertalmio," "CNN-NSIR," and "f-Inpaint" [25]).

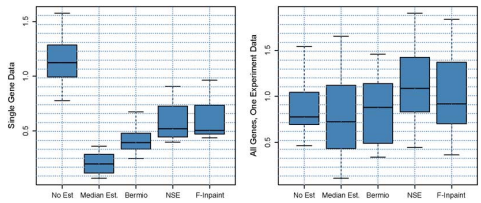

Fig. 4. (a) Box plot for one specific gene. (b) Box plot for one experiment, one image, and data (all genes). (The methods are "No Background Correction," "Median Background Estimation," "Bertalmio," "CNN-NSIR," and "f-Inpaint" [25]).

the robustness of the implemented methodology. In our analysis, every signal less than 100 is considered to be a bad reading and consequently omitted from the analysis.

Fig. 4 gives more detailed plots about the data. Fig. 4(a) shows that not only the range of the signal has been changed due to the reconstruction algorithm application but the average values have also been changed as well. The distribution of the signals, i.e., after applying the reconstruction methods (median background estimation, Bertalmio's method, f-Inpaint, and CNN-NSIR), has a narrower normal distribution with the data skewed toward the lower side. Note that 50% of the genes' values fall within a 0.1 range around 0.5 for our method. Furthermore, while the middle 50% values of "No Est." fall into the 0.4 range width and the whole range is 1.1, the middle 50% values of "CNN-NSIR" fall into the 0.7 range width, and the whole range is 1.6 [see Fig. 4(b)].

V. Conclusion

In this paper, we have presented a novel image-reconstruction framework that attempts to improve the quantification results of the microarray image. Specifically, the framework consists of several components that process a microarray image based on a given mask (could be the output of any automated segmentation process) without human intervention. The algorithm of CNN image reconstruction (i.e., CNN-NSIR), which is outlined in algorithm 1, has been found to have the following advantages over current implementations: 1) the proposed algorithm achieves the reconstruction of the microarray in a simple yet robust way; 2) the algorithm is an operator-free method that takes only the raw data and a mask as the input; 3) the algorithm can be applied on CNN-UM [20] and therefore allows the researchers to process the image itself and get the quantitative data for further analysis not only in efficient time but also with remarkably high accuracy; and 4) as it has been mentioned earlier, the potential of the algorithm

is proven based on the direct comparisons between our proposed approach with other methods such as Bertalmio's, "f-Inpaint," and median-background-estimation methods.

REFERENCES

[1] Anonymous, GenePix 4000 A User's Guide 1999.

[2] P. Arena, L. Fortuna, and L. Occhipinti, "A CNN algorithm for real time analysis of DNA microarrays," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 3, pp. 335–340, Mar. 2002.

[3] D. Bassett, M. Eisen, and M. Boguski, "Gene expression informatics—It's all in your mine," *Nat. Genetics*, vol. 21, no. 1 Suppl, pp. 51–55, Jan. 1999.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. 27th Annu. Conf. Comput. Graph. Interactive Tech.*, New York, 2000, pp. 417–424.

[5] A. Bertozzi and M. Bertalmio, "Navier–Stokes fluid dynamics and image and video inpainting," in *Proc. IEEE Comput. Soc. Conf. CVPR*, 2001, vol. 1, pp. 355–362.

[6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[7] Y. Chen, E. Dougherty, and M. Bittner, "Ratio-based decisions and the quantitative analysis of cDNA microarray images," *J. Biomed. Opt.*, vol. 2, no. 4, pp. 364–374, Oct. 1997.

[8] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing: Foundation and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[9] L. O. Chua and L. Yang, "Cellular neural networks: Theory and applications," *IEEE Trans. Circuits Syst.*, vol. 35, no. 10, pp. 1273–1290, Oct. 1988.

[10] M. B. Eisen, ScanAlayse Mar. 2010 [Online]. Available: http://rana.lbl.gov/eisensoftware.htm

[11] K. Fraser, Z. Wang, and X. Liu, *Microarray Image Analysis*. New York: Taylor & Francis, Feb. 2010.

[12] E. L. Korn, J. K. Habermann, M. B. Upender, T. Ried, and L. M. McShane, "Objective method of comparing DNA microarray image analysis systems," *Biotechniques*, vol. 36, no. 6, pp. 960–967, Jun. 2004.

[13] T. Kozek and T. Roska, "A double time-scale CNN for solving two-dimensional navier—Stokes equation," *Int. J. Circuit Theory Appl.*, vol. 24, no. 1, pp. 49–55, 1996.

[14] A. Lehmussola, P. Ruusuvuori, and O. Yli-Harja, "Evaluating the performance of microarray segmentation algorithms," *Bioinformatics*, vol. 22, no. 23, pp. 2910–2917, Dec. 2006.

[15] P. Luchini, "Two-dimensional numerical integration using a square mesh," *Comput. Phys. Commun.*, vol. 31, no. 4, pp. 303–310, Mar. 1984.

[16] M. M. Oliveira, R. McKenna, B. Bowen, and Y.-S. Chang, "Fast digital image inpainting," in *Proc. Int. Conf. VIIP*, Marbella, Spain, 2001, pp. 261–266.

[17] P. O'Neill, G. Magoulas, and X. Liu, "Improved processing of microarray data using image reconstruction techniques," *IEEE Trans. Nanobiosci.*, vol. 2, no. 4, pp. 176–183, 2003.

[18] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[19] C. Rekeczky, T. Roska, and A. Ushida, "CNN-based difference-controlled adaptive non-linear image filters," *Int. J. Circuit Theory Appl.*, vol. 26, no. 4, pp. 375–423, 1998.

[20] T. Roska and L. O. Chua, "The CNN universal machine: An analogic array computer," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 40, no. 3, pp. 163–173, Mar. 1993.

[21] H. Samartzidou, L. Turner, T. Houts, M. Frorne, J. Worley, and H. Albertsen, "Lucidea microarray ScoreCard: An integrated analysis tool for microarray experiments," *Life Sci. News*, vol. 7, no. 13, pp. 1–10, 2001.

[22] S. Samavi, S. Shirani, N. Karimi, and M. Deen, "A pipeline architecture for processing of DNA microarrays images," *J. VLSI Signal Process.*, vol. 38, no. 3, pp. 287–297, Nov. 2004.

[23] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, no. 5235, pp. 467–470, Oct. 1995.

[24] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed, "Comparison of methods for image analysis on cDNA microarray data," *J. Comput. Graph. Statist.*, vol. 11, no. 1, pp. 108–136, 2002.

[25] B. Zineddin, Z. Wang, and X. Liu, "A cellular neural network for microarray image reconstruction," in *Proc. 16th Int. Conf. Autom. Comput.*, Birmingham, U.K., 2010, pp. 106–111.

Supervised Ordering in $\mathbb{R}^p$: Application to Morphological Processing of Hyperspectral Images

Santiago Velasco-Forero, *Student Member, IEEE*, and Jesus Angulo

***Abstract*—A novel approach for vector ordering is introduced in this paper. The generic framework is based on a supervised learning formulation which leads to reduced orderings. A training set for the background and another training set for the foreground are needed as well as a supervised method to construct the ordering mapping. Two particular cases of learning techniques are considered in detail: 1) kriging-based vector ordering and 2) support vector machines-based vector ordering. These supervised orderings may then be used for the extension of mathematical morphology to vector images. In particular, in this paper, we focus on the application of morphological processing to hyperspectral images, illustrating the performance with practical examples.**

***Index Terms*—Hyperspectral imagery, learning an ordering, mathematical morphology, supervised learning.**

I. INTRODUCTION

Mathematical morphology is a nonlinear image processing methodology based on the application of lattice theory to spatial structures [1], [2]. This means that the definition of morphological operators needs a complete lattice structure, i.e., the possibility of defining an ordering relationship between the points to be processed. From a theoretical viewpoint, a partial ordering is sufficient to construct complete lattices; however, as discussed below, in practical algorithms, we should require a total ordering (i.e., any pair of unequal points must be ordered). Extending ordering to multivariate data is not straightforward, because there is no notion of natural ordering in a vector space, as opposed to 1-D (scalar) case [3]. Therefore, the extension of mathematical morphology to vector spaces, for instance to hyperspectral images, is neither direct nor trivial due to the high-dimensional nature of the data. For a general account on mathematical morphology, the interested reader should refer to [4]–[7], whereas the theoretical formulation of vector morphology is extensively discussed in [8]–[10]. To overcome the lack of natural ordering, the following four families of ordering for multichannel samples have been identified in the literature [3], [11]. In *marginal ordering* (M-ordering), components of vectors are ordered independently (pointwise ordering). This approach produces new vectors which were not originally present in the input image thus, in the case for instance of color images, introducing color artifacts into the output image [8]; this is known in the literature as the *false color problem*. However, by an appropriate color representation, typically a hue/saturation/luminance, marginal ordering can lead to good results [12]. To strictly preserve input vectors, the *conditional ordering* (C-ordering) approach, also known as lexicographic ordering, is frequently used. The C-ordering is based on the ordering of the components selected sequentially according to different conditions or priorities. When all the components are used, the C-ordering is a total ordering. Note that this approach does not use simultaneously the full vector nature of the input. The *P-ordering* is based on the partition of the vectors into groups,

Manuscript received December 08, 2010; revised March 25, 2011; accepted April 07, 2011. Date of publication April 21, 2011; date of current version October 19, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maya Gupta.

The authors are with the Centre de Morphologie Mathematique de l'Ecole des Mines de Paris, France.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2144611