2013 Data Compression Conference

A distortion metric for the lossy compression of DNA microarray images

Miguel Hernández-Cabronero[†], Victor Sanchez[*], Michael W. Marcellin[‡] and Joan Serra-Sagristà[†]

[†] Department of Information and Communications Engineering,
Universitat Autònoma de Barcelona, Barcelona, Spain.

[*] Department of Computer Science,
University of Warwick, Coventry, UK.

[‡] Department of Electrical and Computer Engineering
University of Arizona, Tucson, AZ, USA.

Abstract

DNA microarrays are state-of-the-art tools in biological and medical research. In this work, we discuss the suitability of lossy compression for DNA microarray images and highlight the necessity for a distortion metric to assess the loss of relevant information. We also propose one possible metric that considers the basic image features employed by most DNA microarray analysis techniques. Experimental results indicate that the proposed metric can identify and differentiate important and unimportant changes in DNA microarray images.

I. Introduction

In this section, we motivate the need for introducing a distortion metric to evaluate the performance of lossy compression on DNA microarray images.

A. DNA microarrays

DNA microarrays are used to analyze the function and regulation of the genes of an organism [1]. They are state-of-the-art tools in biological and medical research, and are employed in many areas ranging from the study of metabolism [2] and evolution [3] to the fight against cancer [4], HIV [5] and malaria [6]. Interest in DNA microarrays has grown in the last few years, and an exponential increase of DNA microarray data has been observed [7].

In a DNA microarray experiment, the genetic expression level of two biological samples is compared. In these two samples –for instance one coming from a healthy tissue and the other coming from a tumoral tissue– the same genes might be expressed with different intensities. Analyzing these differences, it is possible to identify genes related to a particular biological process. A DNA microarray experiment consists of several steps, schematically summarized in Figure 1. The biological samples are first dyed with fluorescent markers, usually Cy3 and Cy5 of the cyanine family, and are then left to react on the surface of a DNA microarray chip (step A in Figure 1). The surface contains thousands of microscopic holes or *spots*, each of which is related to an individual gene. This surface is then washed so that each of the biological samples appears only inside the different spots. After that, the DNA microarray chip is scanned using two lasers, each exciting only one of the fluorescent markers, so that one 16 bpp monochrome image is produced for each of the two biological samples (step B in Figure 1). These two monochrome images are usually known as the green and red channels due to the color of the laser needed to excite the fluorescent dyes. The image intensity with which each spot is acquired is proportional to the amount of

1068-0314/13 $26.00 © 2013 IEEE
DOI 10.1109/DCC.2013.26

biological sample that is contained in the spot; that amount is also proportional to the expression intensity of the corresponding gene. Finally, these two images are analyzed to measure genetic expression intensity differences, which are then processed statistically to identify relevant genetic expression alterations (step C in Figure 1).

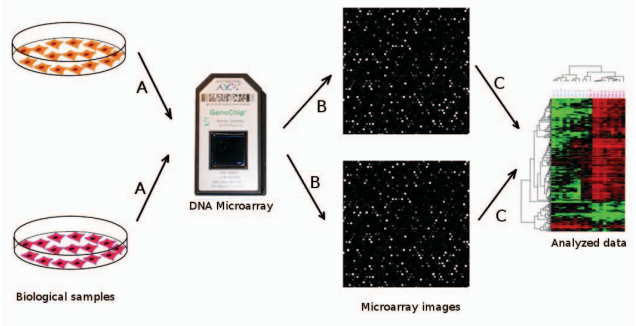


Figure 1: A DNA microarray experiment. The biological samples are first put on a microarray chip, then the chip is scanned to produce two microarray images and finally the images are processed to extract the genetic data.

B. DNA microarray image analysis

DNA microarray images are an intermediate product of DNA microarray experiments: image analysis is performed on these images to extract information about the genetic expression intensity. Unfortunately, these image analysis techniques are not fully mature or universally accepted [8], so they are likely to change in the future. As new image analysis techniques are developed, it will be highly desirable to reanalyze the images to obtain more accurate genetic data. However, in such cases, repeating the whole experiment is not an option because the needed biological samples are usually not available. For this reason, it is important to store the DNA microarray images along with the extracted genetic data.

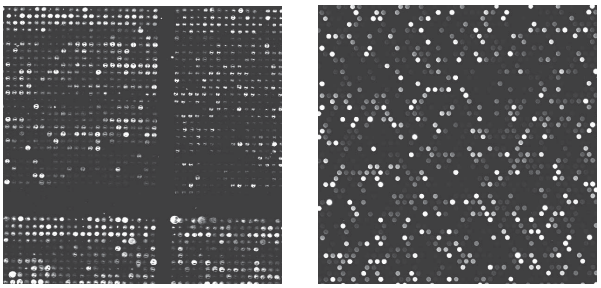
In spite of the increasing reproducibility of DNA microarray experiments, some variability is always present. When samples from the same two tissues are used in different experiments, the produced images, and thus the extracted genetic data, are not identical [9]. In addition, modern DNA microarray chips make use of biological replication, that is, they contain several spots associated to a single gene; even though theoretically these spots should express the same degree of gene activity, in practice they do not [8].

C. DNA microarray image lossless compression

In a DNA microarray experiment, two images –known as the green and red channels– are produced. Nowadays, these images easily exceed 4000×13000 pixels in size, with 16 bits per pixel (bpp) per channel, so that raw file sizes over 100 Megabytes per image are common. Besides, when a DNA microarray study is carried out, several DNA microarray experiments are performed. Thus, a considerable amount of data are produced in laboratories around the world. It is therefore paramount to design efficient storage and transmission methods for this type of images, and data compression comes as the best suited approach to this problem.

DNA microarray images possess several characteristics that make compression a challenging task. Figure 2 shows part of two example DNA microarray images. It can be observed that thousands of irregular round spots of varying intensities are displayed on a dark background. These abrupt and irregular intensity

changes produce high frequencies in the image which are hard to code or predict. Furthermore, the 16 bpp needed to represent DNA microarray images increase the data entropy due to the larger amount of possible pixel values. Altogether, these properties make the compression of DNA microarray images a challenging task, in particular when using lossless compression methods.



(a) (b)

Figure 2: Example DNA microarray images: 600×600 crops with different spot layouts. a) *array3* image from the MicroZip set with square grid spot layout; b) *slide_1-red* from the Arizona set with hexagonal grid spot layout. Gamma levels have been adjusted for visualization purposes.

Table I: Image sets used for benchmarking in the literature. All original images are 16 bpp.

Image set		Images	Size (pixels)
MicroZip	[10]	3	$> 1800\times1900$
Yeast	[11]	109	1024×1024
ApoA1	[12]	32	1044×1041
ISREC	[13]	14	1000×1000
Stanford	[14]	20	$> 2000\times2000$
Arizona	[15]	6	4400×13800

Several authors have proposed compression techniques for DNA microarray images in the last few years. A detailed review of the state of the art in the compression of DNA microarray images can be found in the literature [16]. Both lossy and lossless techniques have been proposed, but lossless techniques are more common. It has been argued that lossless compression is more suitable for DNA microarray images because it guarantees that no relevant information will be missing when reanalyzing the decompressed version of the images [17]. However, this data fidelity is obtained at the expense of poor compression performance results. Table I displays basic information about the sets of images used for benchmarking while Table II reports the lossless compression results yielded by both standard and the best performing microarray-specific techniques.

It can be observed that even the best-performing techniques specific for DNA microarray images are only approximately 1 bpp better than the best standard compression techniques. Moreover, compression ratios equal or better than 2:1 are only found for two of the datasets. Considering these lossless compression results, schemes with higher compression performance are needed.

D. Distortion metrics for DNA microarray image lossy compression

Lossy compression schemes can yield very good compression ratios and as long as the distortion introduced by the lossy compression of DNA microarray images falls below the variability of DNA

Table II: Average lossless compression results for different standard and microarray-specific techniques for the benchmark image sets. All values are expressed in bits per pixel and all images are 16 bpp. Bitrates for non-standard compression techniques are shown as reported by the original authors when data are available for a set. Results for JPEG2000 have been obtained with the best number of wavelet decomposition levels for each set [16].

Algorithm	**MicroZip**	**Yeast**	**ApoA1**	**ISREC**	**Stanford**	**Arizona**
Standard techniques						
Bzip2	9.394	6.075	11.067	10.921	7.867	8.944
JBIG	9.747	6.888	10.852	10.925	7.776	8.858
JPEG-LS	9.441	8.580	10.608	11.145	7.571	8.646
JPEG2000	9.508	6.863	11.050	10.930	8.007	9.099
JPEG2000 + HST [18]	9.157	5.911	10.786	10.624	7.685	8.795
Microarray-specific techniques						
MicroZip [19], 2004	9.843	—	—	—	—	—
PPAM [20], 2005	9.587	6.601	—	—	—	—
Neves [21], 2006	8.840	—	10.280	10.199	—	—
Neekabadi [22], 2007	8.856	—	10.250	10.202	—	—
Neves [17], 2009	8.619	—	10.194	10.158	—	—
Battiato [23], 2009	8.369	—	9.520	9.490	—	—

microarray experiments (Subsection I-B), lossy coding can be regarded as an alternative to lossless coding.

An important number of both standard [24], [25] and microarray-specific [9], [19], [26], [27] lossy or lossy-to-lossless compression techniques have been proposed and discussed in the literature. For lossy coding, it is necessary to assess the loss of relevant information by employing a distortion metric, since it can affect current and future DNA microarray analysis techniques.

To date, only full reference metrics –which need the original image to be calculated– have been considered. Existing metrics based on pixel-wise errors or visual fidelity are not suitable for this purpose [9] because they do not identify changes that could affect the subsequent analysis process. Metrics using pixel-wise errors like MSE, PSNR or SNR consider every pixel in the image equally important [28], hence they are unable to distinguish changes that significantly distort several spots from those that slightly affect unimportant parts of the background. Metrics based on visual fidelity like SSIM or CW-SSIM are designed to identify changes that can be detected by the human visual system [28]; however, DNA microarray images are always computer analyzed –most usually without human intervention– so changes applied to a spot can be difficult to recognize by the naked eye, but still affect greatly the analysis process.

At least two publications have addressed the problem of measuring the information loss due to the compression of DNA microarray images. In [25], the authors propose several methods to evaluate the image distortion based on comparing the results of different classification algorithms like linear discriminant analysis when applied to the extracted mean intensity of each spot. Their methods rely completely on the results of the selected image analysis and classification algorithms and do not explicitly consider image properties used in most analysis algorithms. For this reason, their results are subject to the particularities of the selected image analysis and data classification algorithms and it is unclear whether they would be applicable when other present and future DNA microarray analysis techniques are employed. Even though the authors provide intensive experimental results, they do not discuss which selection of intensity extraction and classification algorithms could be the best as a distortion metric. In [9], a similar approach

is used. Two different image analysis techniques are used on original and distorted DNA microarray images to calculate the spot intensities. After that, simple functions of the extracted intensities are plotted. These functions include the logarithm of the quotient of the intensities extracted for one spot in each of the two DNA microarray images, which is a measure of gene activity. The plots are analyzed to assess the variability using different intensity extraction and compression algorithms. As in [25], the authors do not explicitly use any commonly used image feature nor propose a way to combine the plotted values in order to construct a distortion metric.

E. Paper structure

The rest of this document is structured as follows. In Section II we describe the DNA microarray analysis pipeline and highlight the image features that are commonly used for this analysis. In Section III we propose a distortion metric that employs these image features to detect relevant changes in DNA microarray images. Finally, in Section IV we provide some concluding remarks.

II. Analysis pipeline of DNA microarrays

In this section, we describe the analysis of DNA microarrays. We also identify which processes are basic for any analysis process and what image features are considered in them.

A. Analysis pipeline

In Section I we briefly described how DNA microarray experiments are carried out, as schematically summarized in Figure 1. We now focus on the analysis pipeline that is performed once the DNA microarray images are obtained. The first two subsections describe how the images are analyzed while the third briefly explains subsequent procedures.

1) Gridding and segmentation: The analysis of DNA microarray images begins by locating where each spot is situated. As shown in Figure 1, spots are arranged following a regular grid which must be identified first. This process is known as *gridding* and can be done either automatically without any prior information [19], or using geometrical information provided by the DNA microarray manufacturer.

Once the grid is known, each spot is confined individually to a rectangular area. The next step is to determine which pixels belong to the spot, and which ones are background. This process is known as *segmentation* and is one of the most active research topics on the analysis of DNA microarray images [8]. Many different approaches have been used for this purpose, and it is possible to find clustering-based [29]–[33], threshold-based [34], graph-based [35] and even wavelet-based [36] proposals in the literature.

2) Intensity extraction: The next step in the analysis pipeline is calculating the expression intensity of each gene for each of the biological samples. As previously explained in Section I, the amount of dyed sample inside one spot is proportional to the intensity of the corresponding gene for that biological sample. For this reason, it is possible to estimate the genetic expression intensity by looking at the average value of pixels inside each spot in the DNA microarray images. This process is performed automatically by DNA microarray image analysis algorithms and is known as *intensity extraction* or *feature extraction*.

Unfortunately, in real DNA microarray experiments there are artifacts that distort the scanned images. When the DNA microarray chip is washed, some residues can remain on the surface. To address this and other artifacts, researchers have proposed several *background-correction* algorithms [37]. In their proposals the average intensity of the local background (the pixels outside but near each spot) is calculated and subtracted from the extracted intensity according to different algorithms. In addition, since the two images produced in a DNA microarray experiment are obtained using two different lasers, it is possible that one

of the two images produced in a DNA microarray experiment is globally brighter than the other. For this reason, some analysis techniques compute the global intensity to identify brightness changes and modify the extracted spot intensities accordingly.

3) Normalization and data analysis: After the spot intensities are extracted from the images associated with each of the two biological samples, they are compared to detect relevant differences in the genetic expression intensities. This is commonly know as *data analysis*. Before that, the intensities need to be further processed in order to remove any systematic variation, due for example to dye bias. This step –known as *normalization*– is an active research topic and several normalization techniques have been proposed [8]. Both data analysis and normalization techniques employ only the extracted spot intensities and do not directly consider the images, so they are out of the scope of this work.

B. Key image properties

Having in mind the processes described in Subsection II-A, it is possible to identify at least three key image features that can greatly affect the outcome of a DNA microarray experiment. Since it is our goal to design a distortion metric for DNA microarray images, we next describe these key features more fully and discuss changes that can most alter experimental outcomes.

1) Spot intensity: The value of the pixels inside spots is the most important DNA microarray image feature because it is employed to extract the genetic expression intensities that are employed in subsequent steps of DNA microarray experiments. Only average pixel values are used for this purpose, so it is more important to maintain these mean values unmodified than to achieve pixel-wise fidelity. If we call μ_R and μ_G to the average intensity of the co-located spot in each of the two image channels of an experiment, then the *mean intensity ratio* is defined as μ_R/μ_G. Later analysis steps [38] consider only the mean intensity ratio of the spots, so relative errors in the average intensity values are much more relevant than absolute errors.

2) Local background intensity: The mean intensity ratios are often corrected by subtracting the average intensity of the local background, that is, the pixels that are close but do not belong to a spot. Even though this subtraction can be done in different ways [37], the mean intensity of the local background is always used. Therefore, it is important to keep these local background mean intensity values unmodified.

3) Global intensity: Some normalization techniques rely on the global intensity of each DNA microarray image, i.e., the sum of all pixel intensities [39]. They are based on the assumption that this global intensity should be the same for the images corresponding to each biological sample. As a consequence, producing a deviation on the global intensity can affect the obtained mean intensity ratios proportionally.

To the best of our knowledge, these three features are the basis of every existing analysis technique to the extent of our knowledge, and they are likely to remain fundamental in future techniques as well.

III. Proposed distortion metric

In Section I, we motivated the compression of DNA microarray images and argued that traditional image distortion metrics like MSE or SSIM are not suitable for DNA microarray images. In this section, we propose a novel distortion metric for this type of imagery.

A. Metric definition

Our main goal is to define a distortion metric that is able to summarize information about the three main image features with primary effect on the analysis process. As will be shown in Subsection III-B, common natural image distortion metrics are not suitable for DNA microarray images because they do not

consider these important image features. Since the main purpose of our metric is to detect image changes that can affect present and future analysis, we have designed it to be as conservative as possible; we have also assumed that a segmentation of the image into individual spots and background is provided as input to the metric. Such segmentation may be produced using automatic methods or geometric information from the manufacturer.

In order to produce a compact output that can be more easily interpreted, we have defined our microarray distortion metric (MDM) after the well-known logarithmic metric PSNR:

$$\mathrm{MDM} \;=\; 10\log_{10}\frac{(\mathrm{max\_val})^2}{\mathrm{ME}}. \tag{1}$$

To calculate the "signal-to-noise ratio" of our metric, the maximum value in the image ($\mathrm{max\_val}$) is used as a measure of the signal, while the noise is estimated using our proposed microarray error (ME). The ME must be sensitive to relevant changes in any of the three main features than can affect the analysis process: the mean intensity ratio, the average intensity of the local background and the global image intensity. As explained in the previous section, relative errors in those features are much more important than absolute errors. To estimate the relative distortion in the mean intensity ratio of one spot, we do not need to process both the red and green channel images. If the average intensity inside one spot is multiplied by a factor q in the distorted image, it can be easily proved that the intensity ratio for that spot in the red and green channel images is multiplied by a factor smaller than $\bar{q} = max(q, 1/q)$. In order to detect changes that can affect analysis results even if they appear in isolated spots, we consider only the maximum value of $\bar{q}$ among all spots. A similar conservative reasoning can be applied to the estimations of the distortion of the local background and global intensity. For our metric, we employ the expressions in (2)-(4) to calculate the distortion of the three key image features:

$$r_{\mathrm{spot}} \;=\; \max(\mathrm{max\_spot\_ratio}, 1/\mathrm{min\_spot\_ratio}), \tag{2}$$

$$r_{\mathrm{localBG}} \;=\; \max(\mathrm{max\_localBG\_ratio}, 1/\mathrm{min\_localBG\_ratio}), \tag{3}$$

$$r_{\mathrm{global}} \;=\; \max(\mathrm{global\_intensity\_ratio}, 1/\mathrm{global\_intensity\_ratio}). \tag{4}$$

Ideally, the MDM should produce high signal-to-noise ratios when the image is not distorted enough to affect the analysis results. When such relevant distortions are introduced in any of the key image features, the MDM should decrease toward 0. To achieve this, the definition of ME is based on $\mathrm{max\_val}$ raised to p, a logistic function of r_{spot}, r_{localBG} and r_{global}:

$$p \;=\; 2/(1 + exp(-\alpha(r_{\mathrm{spot}} + r_{\mathrm{localBG}} + r_{\mathrm{global}} - 3))), \tag{5}$$

$$\mathrm{ME} \;=\; (\mathrm{max\_val})^p - \mathrm{max\_val} + \min(\mathrm{max\_val}, \mathrm{MSE}_{\mathrm{image}}). \tag{6}$$

As is obvious from (5), when r_{spot}, r_{localBG} and r_{global} are all close to 1 –a scenario of irrelevant or nonexistent distortions–, p is also close to 1. When distortions relevant to the analysis process are introduced, r_{spot}, r_{localBG} and r_{global} are increased and p approaches 2. In consequence, the first term of (6) varies between $\mathrm{max\_val}$ and $\mathrm{max\_val}^2$. The other terms in (6) are employed for normalization purposes so that the MDM outputs only meaningful values and is able to differentiate lossless and lossy compression by employing the global MSE of the image. The sensitivity of the MDM to changes in the three key image features can be adjusted through the arbitrary parameter α in (5), which controls the speed in which the signal-to-noise ratio is degraded. In our experiments, we have found $\alpha = 3$ to be a balanced choice. When smaller values of α are chosen, the MDM decreases too slowly when essential parts of the DNA microarray images are modified. When larger values are selected, the MDM is too sensitive and produces values close to 0 dB when the key image properties are only slightly modified.

Further experimentation using real compression and analysis scenarios is necessary to elucidate optimal values of α for realistic operating ranges.

B. Experimental results

In this subsection, we illustrate the behavior of our proposed metric and provide further evidence of its suitability for the assessment of the information loss in DNA microarray images.

In our experiments, we have distorted three images from each set shown in Table I in three different manners: we have modified the pixels inside each spot, the pixels inside the local background of each spot and the pixels not inside any spot or local background. In this experiment, we have not modified the images by applying lossy compression so that changes in the images can be more easily located and understood. To identify spot and local background areas, we have employed a Matlab implementation of the circular Hough transform [40], whose results have been further refined to obtain an accurate list of circle centers and radii describing the spots. These results are used to calculate the three types of distortions as well as the segmentation required as input to our proposed MDM. Figure 3a shows the results for the MDM and the PSNR metrics when pixels inside all spots are multiplied by different coefficients. For each spot, we define the *spot ratio distortion* as μ_d/μ_o, where μ_d and μ_o are the mean intensity for that spot in the distorted and original images, respectively. Because of the definition in (2), spot ratio distortions in $(0, 1)$ are substituted by their inverse without loss of generality. In this figure, the horizontal axis represents the average spot ratio distortion. In Figure 3b, we show the results for MDM and PSNR when only the pixels inside local background of each spot are modified. In this figure, the horizontal axis represents the average local background ratio distortion. In Figure 3c, we show the results for applying zero-mean additive white Gaussian noise to pixels outside spots and local backgrounds as a function of noise variance. All results shown in Figure 3 are for the ApoA1 set only. Results for other image sets are similar.

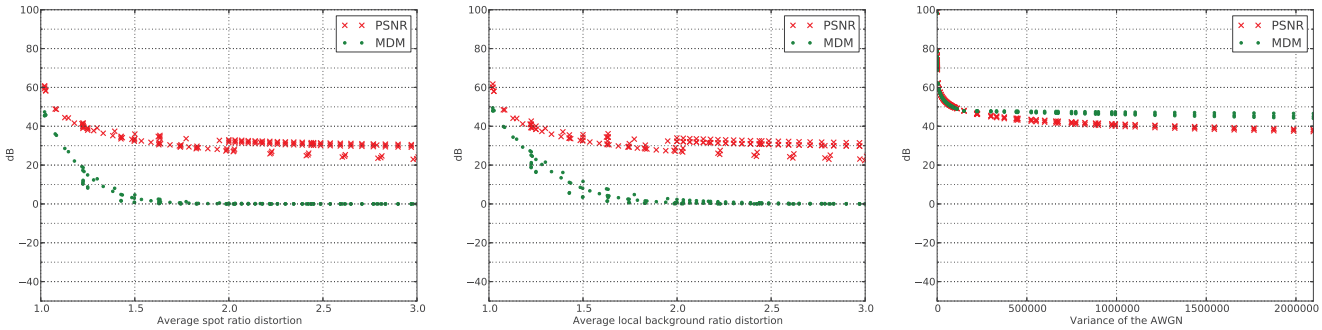
It can be observed that when features that are relevant to the DNA microarray analysis process are modified (Figures 3a and 3b), the MDM decreases rapidly toward 0 dB. The slope in which the MDM decreases when spot ratios are modified can be controlled by the α parameter in (5). Comparing Figures 3a and 3b, it can be seen that the slope is steeper in the former. This is due to the fact that there are more pixels inside spots than inside local backgrounds and they have larger average intensity, so r_{spot} in (5) grows faster. When unimportant changes are applied to the images (Figure 3c), the MDM decreases very slowly. In this case, if the global intensity of the image is not modified, the MDM remains constant at approximately 48 dB, independently of the actual MSE of the image.

These results suggest that the proposed MDM is able to detect changes in DNA microarray images that affect image analysis, whereas unimportant changes do not affect the output of the MDM.

IV. Conclusions and future work

DNA microarray images are broadly employed in biological and medical research to analyze the function of the genes of many different organisms. Large image file sizes motivate the use of coding techniques to help with storage and transmission, but lossless compression has not proved to be effective. On the other hand, lossy compression can provide more compression, but it is necessary to assess whether present and future analysis techniques are affected by the information loss. Distortion metrics like PSNR or SSIM are not suitable for this purpose, so microarray-specific metrics are needed.

The analysis pipeline of DNA microarrays has been discussed, and three key image features have been identified that are the foundation of most current analysis techniques, and very likely for future techniques as well. The identified features are the mean intensity of each spot, of each local background, and the overall intensity of the image. Based on these features, one possible microarray-specific metric has been



(a) (b) (c)

Figure 3: Distortion results for PSNR and our proposed MDM, when applied to three sample images from the ApoA1 set. a) Results after modifying pixels inside spots; b) Results after modifying pixels inside the spot local backgrounds; c) Results after applying additive white Gaussian noise to pixels outside spots and local backgrounds.

proposed, and evidence of its suitability to assess the information loss of DNA microarray images has been provided.

Our future work involves assessing the suitability of the proposed metric to detect changes in the output of standard DNA microarray analysis techniques when different types of distortion are introduced, including the distortion produced by lossy compression.

ACKNOWLEDGEMENTS

MicroZip corpus was kindly provided by Neves and Pinho from the University of Aveiro. The Arizona image set was provided by David Galbraith and Megan Sweeney from the University of Arizona. This work has been partially funded by the Spanish Ministry of Economy and Competitiveness (MINECO) and the Catalan Government under projects FPU AP2010-0172, TIN2009-14426-C02-01, TIN2012-38102-C03-03 (LIFE-VISION) and 2009-SGR-1224.

REFERENCES

[1] S. Moore, "Making chips to probe genes," *IEEE SPECTRUM*, vol. 38, no. 3, pp. 54–60, MAR 2001.

[2] M. Zaparty, A. Zaigler, C. Stamme, J. Soppa, R. Hensel, and B. Siebers, "DNA Microarray Analysis of Central Carbohydrate Metabolism: Glycolytic/Gluconeogenic Carbon Switch in the Hyperthermophilic Crenarchaeum Thermoproteus tenax," *Journal of Bacteriology*, vol. 190, no. 6, pp. 2231–2238, March 15, 2008.

[3] H. Maughan, C. W. Birky, and W. L. Nicholson, "Transcriptome divergence and the loss of plasticity in bacillus subtilis after 6,000 generations of evolution under relaxed selection for sporulation," *Journal of Bacteriology*, vol. 191, no. 1, pp. 428–433, January 1, 2009.

[4] S. Satih, N. Chalabi, N. Rabiau, R. Bosviel, L. Fontana, Y.-J. Bignon, and D. J. Bernard-Gallon, "Gene Expression Profiling of Breast Cancer Cell Lines in Response to Soy Isoflavones Using a Pangenomic Microarray Approach," *OMICS-A JOURNAL OF INTEGRATIVE BIOLOGY*, vol. 14, no. 3, pp. 231–238, JUN 2010.

[5] M. S. Giri, M. Nebozhyn, L. Showe, and L. J. Montaner, "Microarray data on gene modulation by HIV-1 in immune cells: 2000-2006," *JOURNAL OF LEUKOCYTE BIOLOGY*, vol. 80, no. 5, pp. 1031–1043, NOV 2006.

[6] Z. Bozdech, S. Mok, and A. P. Gupta, "DNA Microarray-Based Genome-Wide Analyses of Plasmodium Parasites," in *Malaria*, ser. Methods in Molecular Biology, R. Ménard, Ed. Humana Press, 2013, vol. 923, pp. 189–211.

[7] D. A. Adjeroh, Y. Zhang, and R. Parthe, "On denoising and compression of DNA microarray images," *Pattern Recognition*, vol. 39, no. 12, pp. 2478–2493, December 2006.

[8] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, "Microarray data analysis: from disarray to consolidation and consensus." *Nature reviews. Genetics*, vol. 7, no. 1, pp. 55–65, Jan. 2006.

[9] R. Jornsten, W. Wang, B. Yu, and K. Ramchandran, "Microarray image compression: Sloco and the effect of information loss," *Signal Processing*, vol. 83, no. 4, pp. 859–869, April 2003.

[10] "MicroZip test image set (http://www.cs.ucr.edu/~yuluo/MicroZip)."
[11] "Stanford Yeast Cell-Cycle Regulation Project (http://genome-www.stanford.edu/cellcycle/data/rawdata/individual.html)."
[12] "ApoA1 experiment data (http://www.stat.berkeley.edu/users/terry/zarray/Html/apodata.html)."
[13] "ISREC image set (http://www.isrec.isb-sib.ch/DEA/module8/P5_chip_image/images/)."
[14] "Stanford Microarray Database public FTP (ftp://smd-ftp.stanford.edu/pub/smd/transfers/Jenny)."
[15] "Arizona test image set (http://deic.uab.es/~mhernandez/materials)."
[16] M. Hernández-Cabronero, I. Blanes, M. W. Marcellin, and J. Serra-Sagristà, "Standard and specific compression techniques for DNA microarray images," *MDPI Algorithms*, vol. 4, pp. 30–49, 2012.
[17] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images using image-dependent finite-context models," *IEEE Transactions on Medical Imaging*, vol. 28, no. 2, pp. 194–201, February 2009.
[18] M. Hernández-Cabronero, J. Muñoz-Gómez, I. Blanes, J. Serra-Sagristà, and M. W. Marcellin, "DNA microarray image coding," in *Proceedings of the IEEE International Data Compression Conference, DCC*, IEEE, Ed., 2012, pp. 32–41.
[19] S. Lonardi and Y. Luo, "Gridding and compression of microarray images," in *Proceedings of the Computational Systems Bioinformatics Conference*. IEEE, 2004, pp. 122–130.
[20] Y. Zhang, R. Parthe, and D. Adjeroh, "Lossless compression of DNA microarray images," in *Proceedings of the IEEE Computational Systems Bioinformatics Conference*, August 2005, pp. 128 – 132.
[21] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images," in *Proceedings of the International Conference on Image Processing, ICIP*. IEEE, 2006, pp. 2505–2508.
[22] A. Neekabadi, S. Samavi, S. A. Razavi, N. Karimi, and S. Shirani, "Lossless microarray image compression using region based predictors," in *Proceedings of the International Conference on Image Processing*, 2007, pp. 349–352.
[23] S. Battiato and F. Rundo, "A bio-inspired CNN with re-indexing engine for lossless dna microarray compression and segmentation," in *Proceedings of the 16th International Conference on Image Processing*, vol. 1-6. IEEE, 2009, pp. 1717–1720.
[24] A. Pinho, A. Paiva, and A. Neves, "On the use of standards for microarray lossless image compression," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 563–566, March 2006.
[25] Q. Xu, J. Hua, Z. Xiong, M. L. Bittner, and E. R. Dougherty, "The effect of microarray image compression on expression-based classification," *Signal Image and Video Processing*, vol. 3, no. 1, pp. 53–61, February 2009.
[26] M. R. N. Avanaki, A. Aber, and R. Ebrahimpour, "Compression of cDNA microarray images based on pure-fractal and wavelet-fractal techniques," *ICGST International Journal on Graphics, Vision and Image Processing, GVIP*, vol. 11, pp. 43–52, March 2011.
[27] T. J. Peters, R. Smolikova-Wachowiak, and M. P. Wachowiak, "Microarray image compression using a variation of singular value decomposition," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1-16. IEEE, 2007, pp. 1176–1179.
[28] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98 –117, jan 2009.
[29] V. Uslan and I. O. Bucak, "Clustering-based Spot Segmentation of cDNA Microarray Images," in *2010 Annual International Conference of the IEEE in Medicine and Biology Society (EMBC)*, 2010, pp. 1828–1831.
[30] ——, "Microarray image segmentation using clustering methods," *Mathematical & Computational Applications*, vol. 15, no. 2, pp. 240–247, August 2010.
[31] Z. Li and G. Weng, "Segmentation of cDNA Microarray Image using Fuzzy c-mean Algorithm and Mathematical Morphology," in *Functional manufacturing technologies and CEEUSRO II*, 2011, pp. 159–162.
[32] S. Battiato, G. Farinella, G. Gallo, and G. Guarnera, "Neurofuzzy segmentation of microarray images," in *Proceedings of the 19th International Conference on Pattern Recognition, ICPR*, Dec 2008, pp. 1–4.
[33] S. Battiato, G. D. Blasi, G. M. Farinella, G. Gallo, and G. C. Guarnera, "Ad-hoc segmentation pipeline for microarray image analysis," in *S&T-SPIE Electronic Imaging*, 2006.
[34] I. Rezaeian and L. Rueda, "Sub-grid and Spot Detection in DNA Microarray Images Using Optimal Multi-level Thresholding," in *Pattern Recognition in Bioinformatics*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, vol. 6282, pp. 277–288.
[35] N. Karimi, S. Samavi, S. Shirani, and P. Behnamfar, "Segmentation of DNA microarray images using an adaptive graph-based method," *IET IMAGE PROCESSING*, vol. 4, no. 1, pp. 19–27, FEB 2010.
[36] E. Athanasiadis, D. Cavouras, D. Glotsos, P. Georgiadis, I. Kalatzis, and G. Nikiforidis, "Segmentation of Complementary DNA Microarray Images by Wavelet-Based Markov Random Field Model," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 1068–1074, nov 2009.
[37] J. Ambroise, B. Bearzatto, A. Robert, B. Govaerts, B. Macq, and J.-L. Gala, "Impact of the spotted microarray preprocessing method on fold-change compression and variance stability," *BMC Bioinformatics*, vol. 12, no. 1, p. 413, 2011.
[38] X. Chen and H. Duan, "A vector-based filtering algorithm for microarray image," in *Proceedings of the International Conference on Complex Medical Engineering*, vol. 1-4. IEEE/ICME, 2007, pp. 794–797.
[39] P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. Hughes, E. Snesrud, N. Lee, and J. Quackenbush, "A concise guide to cDNA microarray analysis," *BIOTECHNIQUES*, vol. 29, no. 3, pp. 548+, SEP 2000.
[40] T. Atherton and D. Kerbyson, "Size invariant circle detection," *Image and Vision Computing*, vol. 17, no. 11, pp. 795–803, sep 1999.