A Fully Automated Gridding Technique for Real Composite cDNA Microarray Images

STEFFY MARIA JOSEPH, (Student Member, IEEE), AND P. S. SATHIDEVI, (Member, IEEE)

Department of Electronics and Communication Engineering, National Institute of Technology, Calicut 673601, India

Corresponding author: Steffy Maria Joseph (steffymaria89@gmail.com)

ABSTRACT Genome-wide screening using microarrays of DNA will be of great use in the early diagnosis of diseases such as cancer and HIV. It also makes use of gene discovery, pharmacogenomics, toxicogenomics, and nutrigenomics for other applications. A DNA microarray image lays out an orderly arranged specific gene regions called spots. Microarray image analysis consists primarily of preprocessing, spot area gridding, spot segmentation, and intensity extraction. The first two phases are focused on this work: preprocessing and gridding. The experiment is conducted on real composite cDNA microarray images. A composite microarray image is formed by suitably stacking a red channel image and a green channel image acquired from a microarray experiment either in the RGB domain or in the GRB domain. The blue channel is kept as zero. In order to reduce the challenging problems of microarray images, an efficient preprocessing algorithm is proposed here for these composite images. We have developed a fully automated gridding algorithm integrating global subgrid gridding and local gridding of spots. This technique extracts the structural information namely inter-subgrid spacing, inter-spot spacing and spot center position to achieve efficient gridding. The traits of a microarray image are evaluated using three parameters namely Mean square error, Naturalness quality image evaluator and degree of contrast. The accuracy of the experimental results indicates that this combined preprocessing and gridding technique performs better than existing competitive methods in SIB, GEO, SMD and DeRisi datasets which are most commonly used by the research community for microarray image analysis techniques.

INDEX TERMS cDNA, composite microarray images, contrast enhancement, denoising, fully automated gridding, genome-wide monitoring, global gridding, hybridized spots and non-hybridized spots, local gridding, spot region extraction.

I. INTRODUCTION

Microarray is a cutting-edge bioinformatics technology of today. cDNA microarray image analysis is an important part of microarray experimental studies that can potentially have a significant impact on subsequent analyzes such as clustering or detection of differential genes [1]. The basics of DNA, cDNA, DNA chips and cDNA microarray image formation are described in our earlier work [2]. A comprehensive virtual laboratory experiment is available in [3], by which we can investigate the differences between healthy cells and cancer cells using DNA microarray.

Gridding is one of the most important steps in microarray image analysis. This step includes Global gridding and Local gridding. Global gridding assigns coordinates to the subgrids or sub-arrays in the entire image whereas local gridding assigns coordinates to each of the hybridized or non-hybridized spots within the sub-arrays. The located spot region can be differentiated into the foreground spot region and the background region using the various segmentation rules. Practically, background intensity around the spot is not zero. A suitable background intensity correction method should be done either globally or locally before the extraction of image intensities. Flagging is a procedure to be conducted for removing or marking poor quality intensity features after the spot segmentation. Finally, the generated intensity values will pave the way for the clustering of differentially expressed genes.

A preprocessing stage will enhance the accuracy of gridding, segmentation, intensity extraction, and quantification by reducing the common artifacts present in images. So this stage is unavoidable. Hence we focused on an automated

method for preprocessing combined with gridding of the images. Automating this part of the procedure enables high-throughput.

Over the past decades, numerous manual gridding, semi-automatic gridding, and fully automatic gridding techniques have been mentioned in the literature. In manual gridding method, all the parameters mainly number of subgrids, number of spots in each subgrid, and spot sizes required for gridding are provided manually [4]–[7]. However, the semi-automatic gridding method demands the user to input some of the gridding parameters in order to limit the preset variables [8], [9]. Automatic gridding algorithms detect all spots of microarray images without any preset parameters or user intervention. Most automated gridding strategies proposed in the literature are for the spot detection's local grid.

Related works areas in this field are as follows. In 2003, authors used morphological operators to provide horizontal and vertical projections of the image in a morphological mathematical model [10]. This method is simple and computationally fast but susceptible to misalignment due to noise and spurious or missing spots. In 2004, a grid alignment technique [11] was implemented, which is data-driven and uses statistical analysis of 1-D projections of image directional edge features. The approach is stable to row and column-spaced spots irregularly, independent of the spot features and reduces the number of free parameters by algorithms powered by data. The proposed 2006 hill-climbing search [12] is intended to locate a local minimum (or maximum) that continually moves in the direction of the objective function's decreasing (or increasing) value. It is necessary to initialize one or more starting pixels that will affect the final result. In 2006, authors calculated the maximum a priori grid estimate [13] for microarray image bayesian gridding. Radon transform properties are used to obtain the positions of the reference grid. The Genetic Algorithm (GA) [14], [15] implemented in 2008 effectively define line segments, which constitute the boundaries in an image between neighboring subgrids or spots. Though no input parameters or user interaction are needed, GA computation is a slowpoke.

In 2010, authors of [16] found spot detection gridding for M^3G. The rotation of images is calculated using the radon transform. The SVM classifier is used to create the full grid line between the spot columns and the spot rows. Optimal multilevel threshold gridding (OMTG) [17] established in 2011, applies radon transform to adapt original images to the rotation. The appropriate number of thresholds is found, and a refinement process is used to improve the performance of the system for efficient global gridding and local gridding. In 2013, the approach [18] proposed automated gridding of microarray images of DNA using an optimum sub-image. The gridding lines are determined using the intensity projection profiles from this sub-image. In 2013, authors used their sub-array gridding unit with a modified Otsu algorithm [19], which provides good accuracy with less computational time. For increasing the precision of spot detection, heuristic rules are applied to the grid lines acquired. Using a background

estimation process, spot pixels are enhanced before gridding. In 2015, the method [20] described a sub-array gridding technique based on intensity projection profiles following a histogram-based threshold enhancement of pixels. For detecting and correcting grid line errors a refinement technique is applied.

In 2016, a generalized methodology is proposed for the spot addressing of microarray images which include hexagonal or rectangular grid patterns [21], [22]. Hybridized spots are segmented based on simple segmentation rules, and an increasing concentrate polygon (GCP) algorithm detects non-hybridized spots. Eventually, it measures a Voronoi diagram to distinguish each position in a single cell. In 2017 the authors of [23] used microarray images to implement multiple image processing techniques for cancer detection. A complete microarray image analysis is done before the disease diagnosis. Microarray image blocks are identified by the computation sum of pixels in a row/column and taking the local minima. The successive local minima in each sub-arrays are calculated for spot detection by the period derived from the auto-correlation function applied on the histogram of each block. In 2017 a shape-independent algorithm [24] for fully-automated gridding of cDNA microarray images is implemented to minimize grid line errors. This technique involves finding the subarray in an image by using the Blackman window length variable, Otsu threshold-based image contrast enhancement, and defining image objects such as spots and noise blocks through the 8-connected labeling procedure. In 2018, a Cross covariance approach [25] with a mean image profile is used to compare the known and unknown locations in microarray images. Noisy pixels are removed by using structural processing methods before the gridding stage. In 2019, authors applied an improved Otsu method which is optimized by multilevel thresholds to achieve precise gridding [26]. Multilevel thresholds are obtained using physical information from source microarray images. In 2019, the authors introduced a fully automatic gridding technique [27], applying multi-resolution analysis to obtain horizontal and vertical image projection profiles in global and local grid units. Upon identification of the sub-array, an adaptive threshold approach is determined to get the grid lines for locations.

Although many attempts are focused on this particular gridding step, they are almost failed to properly gridding tough images (high-density noisy images, images with a large number of missing spots (non-hybridized spots), low-resolution spots, almost same measurements for subgrid and spot spacing) in the databases. Images in the same databases have a different number of sub-arrays and spots. Sub-array spacing, spot spacing, size and shape of spots are also dissimilar. Images feature with almost the same inter sub-array spacing and inter spot spacing can cause an error in global gridding lines. Many of the contrast enhancement methods in the literature boost the noisy pixels along with the spot pixels. Hence, the accuracy of global and local gridding of spots is observed very poorly in high-density noisy images.

Also, some researchers have not used universal microarray databases namely GEO, Derisi, SMD available for developing and testing their methods. Therefore, the performance of their methods in automated gridding cannot be compared.

Our work employs a novel preprocessing method which gives a good contrast with reduced noise effect on high-density noisy images. The preprocessing work is done on composite images, which combines the advantages of three-dimensional (3D) median filtering and the contrast enhancement by dehazing technique. This dehazing technique works well in poor contrast images. The contrast-enhanced image is top-hat filtered to reduce high fluorescent block noise to a large extent. The size of the structuring element is automatically derived from the obtained spot spacing measurement. The proposed fully automated gridding method exploits the structural information (inter subgrid spacing, inter spot spacing, and location of spot centers) from the images so that the algorithm can be automated for different featured microarray databases. The grid lines are calculated in such a way that spots are located in the middle region of their grid compartments (Perfectly gridded). This lessens the error associated with marginally gridded (more than 80% of spot pixels reside in the equivalent compartment) and incorrectly gridded (less than 80% of spot pixels reside in the equivalent compartment) spots. Also, three quality parameters are used to analyze the traits of images which aid the gridding accuracy of perfectly gridded spots in tough images.

The rest of this paper is structured in six sections. Microarray images, their ideal characteristics, and the challenges associated with practical/real images are described with figures in section II. Section III mentions the details of real microarray image corpora used in our methods. Methods and detailed algorithms of preprocessing, global gridding, and local gridding are well explained in section IV. The results of the proposed work and comparison with most competitive works are discussed in section V, while the conclusion is written in section VI.

II. MICROARRAY IMAGE

The microarray experiment's outputs are two high-resolution 16-bit grayscale images in TIFF format, for both the green and red channels. Each image contains regularly arranged unique gene probes called spots. A relative level of gene expression is given by the ratio of gray-level intensities of the red spot to the green spot.

A. CHARACTERISTICS OF AN IDEAL MICROARRAY IMAGE
The ideal microarray image has the following features:

(1) All the sub-arrays have the same number of rows and columns, (2) There is no rotation or inclination of spots in horizontal as well as in the vertical direction, (3) Inter sub-array spacing and inter spot spacing are regular, (4) Sizes and shapes are the same for all the spots, (5) No dust and other contamination are affected on the microarray slide, (6) A uniform background intensity should remain across the image, (7) Spots should reflect only the true measures of fluorescence intensities of a corresponding dye of interest.

B. CHALLENGES IN REAL MICROARRAY IMAGES
This part addresses the main challenges of real microarray images with illustrations.

Real microarray images exhibit a large dynamic range. It requires 16 bits/pixels whereas 8 bits/pixels are used in natural images. The size of the image/channel varies approximately from 2MB to 120MB which directly depends upon the total number of subgrids and spots immobilized in the DNA chip. Images are of poor contrast and possess variable-sized subgrids. All the spots are not clearly detected. This is demonstrated in Fig. 1a. Also, it is visible that the second subgrid size is 5×7 and other subgrids have the dimension of 7×7 in the same image. Spots with variable sizes and shapes are shown in Fig. 1b. All are not in circular in shape. The histogram comparison of a natural image, a graphics image, and a microarray image [28] is displayed in Fig. 1c. The pixels in the microarray image use only a smaller fraction of all possible intensities. Also, the prevalence of low-intensity spots pixels with a large background gives a unimodal distribution. Common image enhancement methods like power-law transformation and histogram equalization will result in a washed-out image and hence they are not directly suitable on these images. The fifth row of the subgrid image

(a) poor contrast, subgrids of different sizes

(b) variability in spot shape and size

(c) unimodal histogram

(d) large number of missing spots/ non-hybridized spots

(e) tilt within subgrids
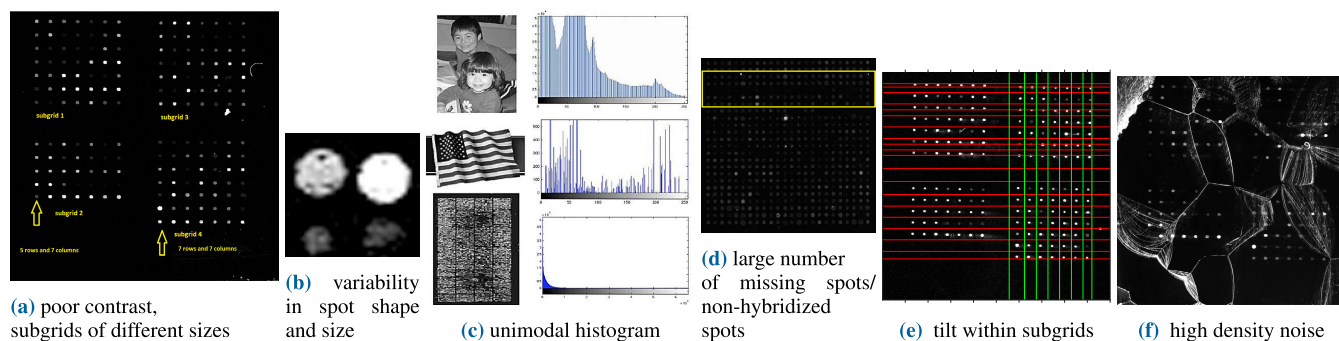
(f) high density noise

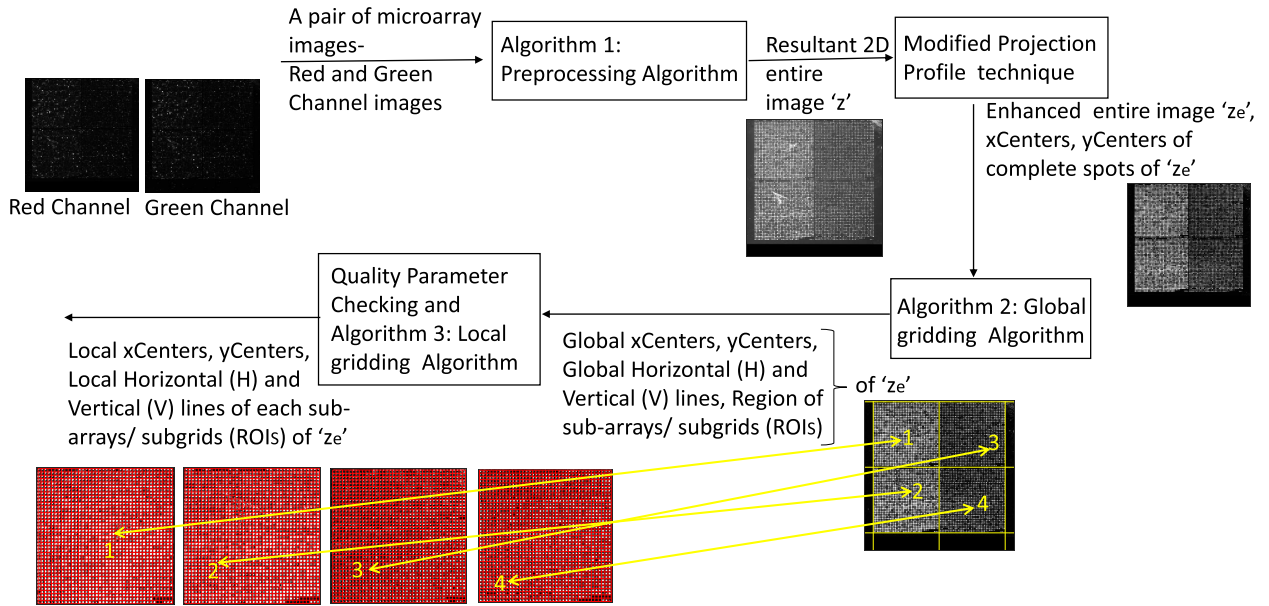**FIGURE 1.** Challenges present in the real microarray databases.

FIGURE 2. Overall block diagram of the proposed fully automated method.

in Fig. 1d contains a large number of missing spots. This will miss the horizontal grid line due to the low value of mean intensity in that row. A large number of missing spots are often visible in real microarray images. Fig. 1e depicts a contrast-enhanced image with reduced noise effect. However, false and redundant horizontal grid lines are appeared due to the tilt or rotation of subgrids or spots within the images. High-density noise is affected like a spider-web on all the subgrids in the image displayed in Fig. 1f. This is the toughest image named 'Def665Cy5.tif' which causes false global gridding of subgrids and local gridding of spots in our experimental databases. None of the works proposed so far in the literature obtained a 100% accuracy with perfect gridding of spots for the aforementioned image. However, the method proposed in this work gives a 100% accuracy. Most of the microarray images experience a non-uniform background intensity around the spots and sometimes contain block noises similar to the size and the intensity level of real spots. This is a big challenge in the right gridding of spot regions.

Hence, a preprocessing algorithm is inevitable to address all the issues before the image analysis for different applications.

III. DATASETS

Four different cDNA microarray databases that are most commonly used in the literature have been selected for evaluating the performance of the proposed method. Databases are two-channel image datasets. Databases and image features are exhibited in Table 1. Images in the datasets possess different scanning resolutions, inter-subgrid spacing, inter-spot spacing, and spots with different features. All the challenges addressed in section II.B are considered to ratify the flexibility of our proposed method.

TABLE 1. Real Microarray image corpora used for evaluating the performance of the proposed technique.

Real Dataset	SIB	Derisi	GEO	SMD
Name	Swiss Institute of Bioinformatics	Joe DeRisi Individual	Gene Expression Omnibus	Stanford Microarray Database
Image Format	Tiff	Tiff	Tiff	Tiff
No. of Images	14	14	8	20
No. of Subgrids/ Image	4	4	48	16
Subgrid Layout	2×2	2×2	12×4	4×4
Spot Layout	one 5×5 and three 7×7	40×40	13×14	17×17
No. of Spots/ Image	182	6400	8736	4624
Image Resolution (approximate)	1000×1000 **2 MB**	1024×1024 **2 MB**	5997×2200 **25 MB**	2164×2200 **8 MB to 13 MB**

The first database includes 14 microarray images from the Computational Cancer Genomics (CCG) group of the Swiss Institute of Bioinformatics (SIB) [29]. The images have been named using Def, followed by experiment IDs 661, 662, 663, 664, 665, 666, 667 and channel number is Cy3 or Cy5. Each image contains four subgrids. Subgrids 1, 3 and 4 contain 49 spots and subgrid 2 contains 35 spots.

The second database of 14 images is drawn from Joe DeRisi's individual tiff files [30]. These images correspond to channels 1 and 2 for experiment IDs OD730, OD046, OD014, OD180, OD690, OD080, and OD370. Each image contains four subgrids and each subgrid contains 1,600 spots.

The third database consists of a set of eight images which are selected from Gene Expression Omnibus (GEO) [31]. Images correspond to channels 1 and 2 for experiment IDs GSM15898, GSM16389, GSM16391, and GSM16101. Each image contains 48 subgrids and each subgrid contains 182 spots.

The fourth database provides a set of images drawn from the Stanford Microarray Database (SMD). Each image contains four subgrids and each subgrid contains 289 spots.

IV. METHODS

The proposed methodology includes three main algorithms: Preprocessing Algorithm for image enhancement, Global gridding algorithm to get sub-array regions and Local gridding algorithm for spot regions which will be discussed in detail on the pages concerned. The overall block diagram of the proposed fully automated method is shown in Fig. 2.

A. PROPOSED PREPROCESSING ALGORITHM FOR COMPOSITE IMAGES

The major steps of the preprocessing algorithm are listed in Algorithm 1. Initially, a composite image is made by suitably stacking as [Red channel, Green channel, Blue Channel] or as [Green channel, Red channel, Blue Channel]. There is no blue channel for microarray images, hence it is kept as zero planes. Exchange of Red plane and Green plane is done based on the mse_r or mse_g values mentioned in steps 1, 2 and 3 of Algorithm 1. This stacking of planes can form a composite image that exploits the advantages of median filtering in the 3D domain. The Signal Processing Toolbox in Matlab computing environment allows the easy computation of 3D median filtering. The obtained median values are position-dependent and relative to the stacking of Red and Green planes. This is done in step 4. The results of 3D median filtering are displayed in Fig. 3. The noise effect is considerably decreased in the GRB domain than in the RGB domain after the 3D median filtering for the image Def665. This is visually justified in Figs. 3 d, and 3 e. The converted image back to the RGB domain after the filtering is depicted in Fig. 3 f. Contrast enhancement is implemented by using the dehazing algorithm [32] implemented for low light images in [33]. Obviously, microarray images are of poor contrast. All the spots in the image are not equally visible and are analogous to a low light image scene contents.

It is noted that the histogram of a low light image in the complemented version is similar to the histogram of a hazy image. Therefore, step 5 in Algorithm 1 is done. The resultant image from step 5 is converted into grayscale in step 6. In step 7, rotation adjustment is done by using radon transform to minimize the tilt within subgrids and spots. These types of rotations are often visible in microarray images within a certain angle range. Radon coefficients for each deviation angle are normalized and entropy is calculated. The best rotation angles are chosen based on the minimum value of entropy. An affine map is created for the anti-rotation which uses the derived best rotation angles in x and y direction to spatially transform the input image. This is explained in [17]. Extra boosting of contrast is again done by sharpening the image in step 8. Step 9 gives a preprocessed image z, which is 3D median filtered, contrast-enhanced, and tilt corrected image in 2D.

Composite images in 3D are generated mainly to exploit the benefits of 3D median filtering and contrast enhancement by haze removing algorithm.

B. MODIFIED PROJECTION PROFILE TECHNIQUE

The projection profile technique endeavors the structural data of microarrays. Microarrays generally hold the regular grid of spots. The mean intensity for each column of the image will give an estimation of the number of spot centers and allows the calculation of inter-spacing between the spots. A horizontal intensity projection profile gives the number of vertical separators or grid lines. Similarly, a vertical intensity projection profile gives the number of horizontal grid lines.

1) STEPS FOR CREATING HORIZONTAL INTENSITY PROFILE FOR VERTICAL SEPARATORS

Horizontal intensity profile is calculated by (1), where $[r, c] = size(z)$. The auto-correlation is taken to enhance the self-similarity by reducing the irregularities in the profile. This is illustrated in Fig. 4 and in Fig. 5.

The estimated period for spot spacing can be found by taking the median value of differences of peaks.

$$xProfile = \frac{1}{r}\sum_{i=1}^{r} z(i,j); (j \in [1, c]) \qquad (1)$$

2) REMOVING BACKGROUND NOISE BY MORPHOLOGICAL TOP-HAT FILTERING METHOD

Top-hat filtering can be exploited to rectify uneven illumination when the background is dark. The structuring element used here is a flat linear structuring element 'SE' with the length of line the same as the estimated period computed. The image 'z' in step 9 of Algorithm 1 is top-hat filtered, and followed by median filtered in order to reduce the effect of high-density block noises generated in the microarray experiment as well as during the image acquisition time. Amplified noise after the contrast enhancement also will be mitigated in this stage. The resultant image can be named
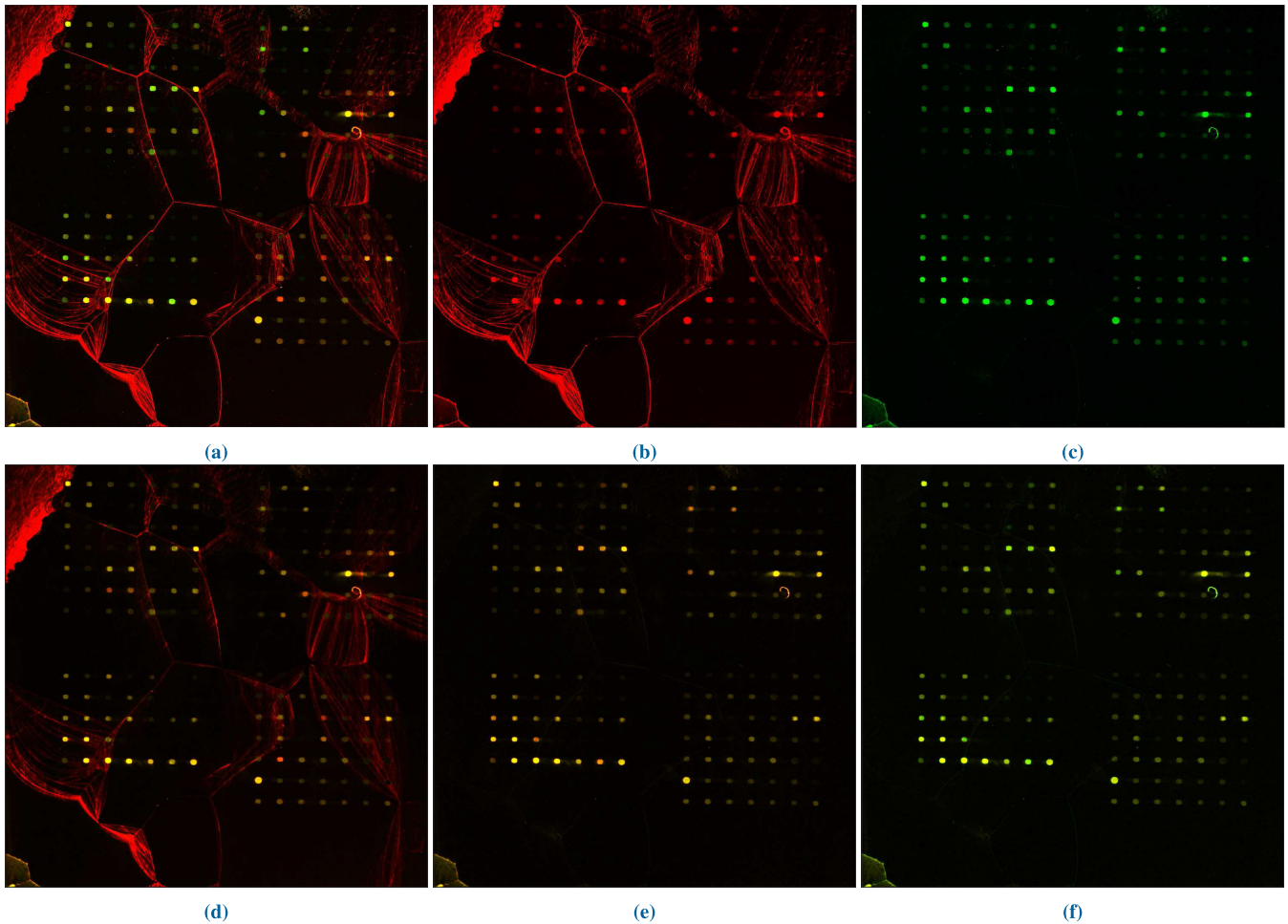
FIGURE 3. Image Def665 from SIB dataset, (a) source image in rgb domain, (b) Red Plane ($mse_r = 316.0161$), (c) Green Plane ($mse_g = 22.1008$), (d) denoised in rgb domain, (e) denoised in grb domain, and (f) denoised image in grb domain is converted back to rgb domain. (In Algorithm 1, the resultant image illustrated in (e) is obtained based on the lower MSE value. The image in (d) is displayed only for the visual comparison with (e).)

as 'z_e'. Fig. 6 displays the resultant figures of z after the pre-processing method, intermediate stage after the morphology operation 'opening', and enhanced z_e of image OD370 from the DeRisi dataset. Comparison of z_e with the source image and the resultant images obtained after the implementation of common enhancement methods mainly histogram equalization (HE) and contrast limited adaptive histogram equalization (CLAHE) are exhibited in Fig. 7. HE works on the entire image and enhances the overall contrast of the image. This method gives a washed-out image in microarray image datasets and hence they are not directly suitable. CLAHE boosts the contrast of small tiles or regions of the image so that the histogram of the output region approximately matches the required histogram. However, this enhancement method is not at all good enough to provide a sufficient contrast level for spots and background in the microarray images. The horizontal intensity profile is again calculated on 'z_e' by (1) and top-hat filtered using the same structuring element 'SE' to obtain an enhanced profile.

3) SEGMENT PEAKS AND LOCATE CENTERS

The enhanced profile is segmented by Otsu's method. Otsu method computes a global threshold level to minimize the intra-class variance of the thresholded white and black pixels. This is a good example of the image processing techniques that are often useful when analyzing 1D data [34]. Each peak region of the binarized image can be numbered and corresponding centroids can be extracted. These centroids conform to horizontal centers '*xCenters*' of the spots. The middle points between adjacent horizontal centers provide vertical grid point locations. Image is transposed and the aforementioned steps are repeated for achieving the vertical intensity profile, *yCenters*, and the horizontal separators. Original and enhanced horizontal and vertical profiles of Def665 are shown in Fig. 8. Enhanced profiles got clean and anchored peaks with regular spot spacings when they are compared to their original profiles.

We illustrated most of the resultant figures of image Def665. This image is the most challenging image in the

---

Algorithm 1: Proposed Preprocessing Algorithm

initialization: $x_1 \leftarrow$ red channel image;
$x_2 \leftarrow$ green channel image
Step 1: Find median filtered images x_{m1} of x_1 & x_{m2} of x_2
Step 2: Check $mse(x_1, x_{m1})$ and $mse(x_2, x_{m2})$;
let $mse_r \leftarrow mse(x_1, x_{m1})$ and $mse_g \leftarrow mse(x_2, x_{m2})$
Step 3: if $(mse_r \leq mse_g)$ **then**

$\quad$ stacking image planes to get composite image;
$\quad\quad x_3(:,:,1) \leftarrow x_1;$
$\quad\quad x_3(:,:,2) \leftarrow x_2;$
$\quad\quad x_3(:,:,3) \leftarrow zero;$
$\quad\quad x_3$ is stacked as RGB;

else

$\quad\quad x_3(:,:,1) \leftarrow x_2;$
$\quad\quad x_3(:,:,2) \leftarrow x_1;$
$\quad\quad x_3(:,:,3) \leftarrow zero;$
$\quad\quad x_3$ is stacked as GRB;

end
Step 4: Do the 3D median filtering on obtained composite image x_3
Step 5: Contrast Enhancement is done by using dehazing algorithm [32] implemented for low light images [33];
(x_3 converted back to RGB domain after Step 4)
Do *complement* of x_3
Apply haze removal
Take *complement* again to get back the resultant image in RGB domain (original domain)
Step 6: Convert resultant image into its grayscale version x_{3g}
Step 7: x_{3g} is tilt corrected using the technique implemented in [17]
Step 8: Find the edge features using a high pass filter on the resultant image and get the sharpened image.
Step 9: Add the resultant image and sharpened image.
Final Preprocessed image z is denoised, tilt corrected and contrast-enhanced image in 2D

SIB dataset which is not globally gridded 100% and locally gridded 100% by anybody to the best of our knowledge.

C. QUALITY PARAMETERS APPLIED TO ANALYZE THE TRAITS OF IMAGES

The following three quality parameters are used to analyze the traits of images in the databases.

1) Mean-Squared Error (MSE): Calculates the mean-squared error (mse_r / mse_g) between red/green channel images and their corresponding median filtered images. It is used in switching to RGB/GRB domain before the effective 3D median filtering.

2) Naturalness image quality evaluator (NIQE): This No-reference image quality score [35] calculates the niqe score with respect to a custom niqe MODEL computed for microarray images using the FITNIQE function in

Matlab. The MODEL contains visually good images in equal proportion from the experimental databases. NIQE value is used to understand the perceptual quality difference of an enhanced image from the source image. The enhanced image is more similar to the source image if the NIQE value of the enhanced image is more close to that of the source image (contrast enhancement in Algorithm 1 for 'z_e is not adequate and need extra boosting).

3) Contrast degree (RMS_C): RMS contrast uses the standard deviation of the image and is computed by (2). RMS_C of the source image (gray-scale) and enhanced image are found. Some enhanced images still cause grid line errors. It is believed that the level of enhancement is not adequate in those images. We will be multiplying a scale of contrast difference ($Cont_{diff}$) to those enhanced images by analyzing their $Niqe_{diff}$ values ($Niqe_{diff} = Niqe_{new} - Niqe_{old}$). If $Niqe_{diff}$ is less than a particular threshold (empirically optimized), that 'z_e' is extra boosted in contrast value.

$$RMS_c = \sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \bar{I})^2} \quad (2)$$

$\bar{I}$ is the average intensity of all pixel values in the image and MN is the product of the number of rows and the number of columns in the image.

The values obtained for each image are listed in Table 2. Also, the best rotation angles computed by radon transform technique in x and y directions to minimize the tilt are added in Table 2.
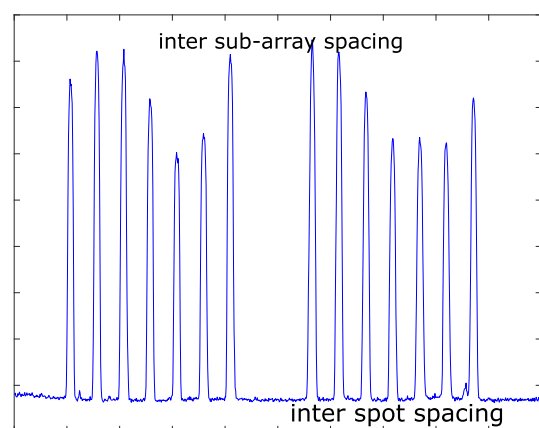
FIGURE 4. Horizontal intensity profile of image 'z'. The image used is Def661 from the SIB dataset.
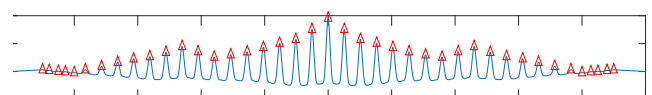
FIGURE 5. Auto-correlation of the horizontal intensity profile in Figure. 4.
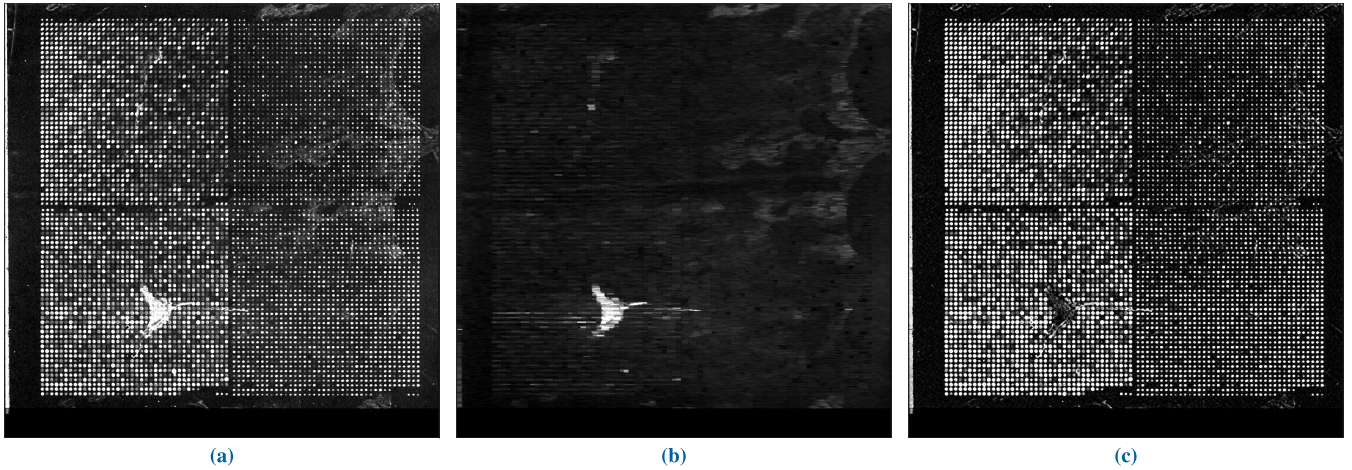
FIGURE 6. Image OD370 from DeRisi dataset used in the proposed preprocessing algorithm, (a) image z after Algorithm 1, (b) image after 'Morphology opening'(intermediate stage), and (c) image z_e after Top-hat filtering and Median filtering.

(a) source image in gray-scale **(b)** after HE **(c)** after CLAHE **(d)** enhanced image z_e

(e) histogram of (a) **(f)** histogram of (b) **(g)** histogram of (c) **(h)** histogram of (d)
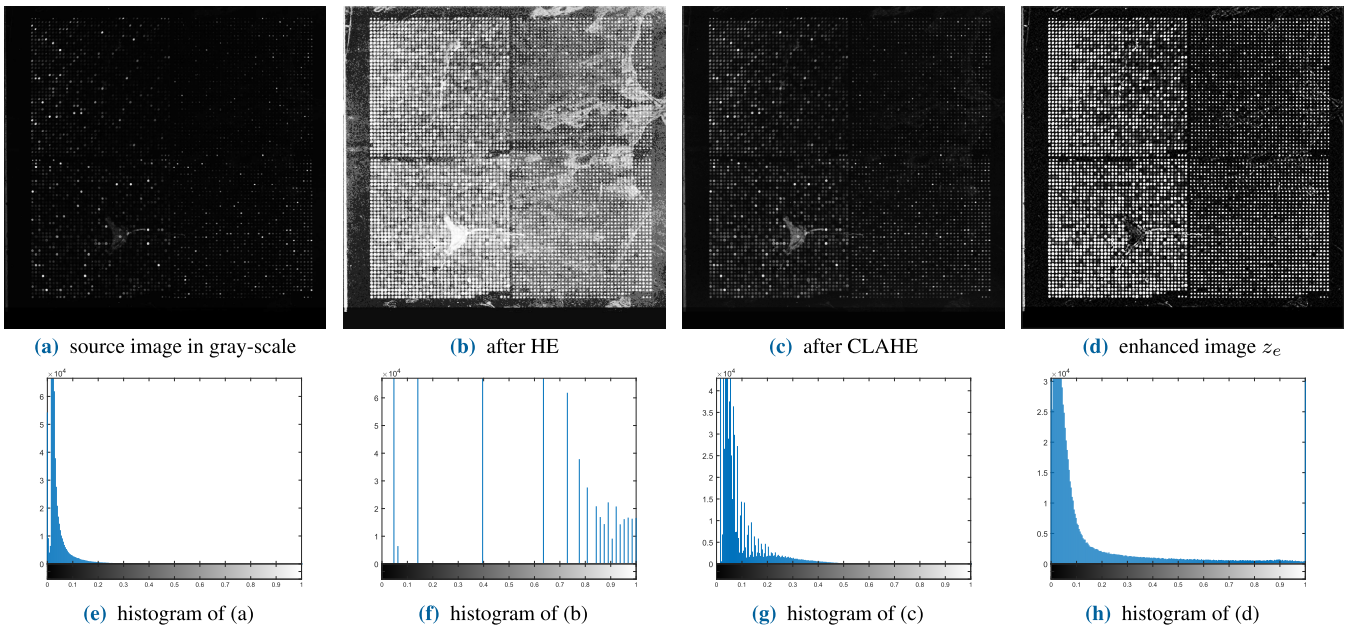
FIGURE 7. Image OD370 from DeRisi dataset, Comparison of resultant images after histogram equalization (HE) and contrast-limited adaptive histogram equalization (CLAHE) with the enhanced image z_e in the proposed method. While comparing the corresponding histograms, the histogram of z_e reveals a better pixel distribution than the other histograms in (e), (f), and (g).

(a) original horizontal profile **(b)** enhanced horizontal profile **(c)** original vertical profile **(d)** enhanced vertical profile
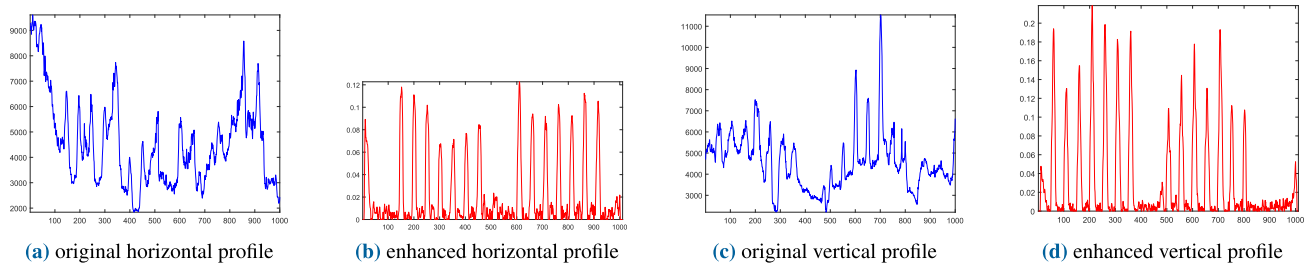
FIGURE 8. Image Def665 from SIB dataset. Resultant horizontal and vertical profiles of the source image (a) and (c) and enhanced image z_e (b) and (d).

D. PROPOSED GLOBAL GRIDDING

Major steps of the Global gridding algorithm are listed in Algorithm 2. Horizontal centers '$xCenters$' obtained from the modified projection profile technique of the enhanced image z_e are loaded initially in the algorithm. Inter spot spacings of $xCenters$ are calculated and $spot\_x$ gets the estimated period of spot spacings in step 1. In projection profiles, noisy $xCenters$ can be seen at the beginning and at the endpoints,

---

Algorithm 2: Proposed Global Gridding Algorithm

Result: Refined Global vertical gridding points

Step 1: Load $xCenters$ from enhanced image z_e;

Let $dist\_x \leftarrow$ differences of $xCenters$;

$spot\_x \leftarrow$ median of $dist\_x$

Step 2: Refining $xCenters$

Let $t\_x1 \leftarrow (spot\_x - 20)$ *and* $t\_x2 \leftarrow (spot\_x + 20)$ thresholds to remove noisy $xCenters$;

Checking beginning 3 points ($i = 1, 2, 3$) and end 3 points ($i = end, end - 1, end - 2$)

if $dist\_x(i) > t\_x2$ *or* $dist\_x(i) < t\_x1$ **then**

    delete $xCenters(i)$;

end

Refine $dist\_x \leftarrow$ differences of new $xCenters$;

Step 3: Calculate an array n_x of distances greater than $spot\_x$;

Let $lx \leftarrow 1$;

for $n \leftarrow 1$ **to** $length(dist\_x)$ **do**

    if $(dist\_x(n) > spot\_x)$ **then**

        $n\_x(lx) \leftarrow (dist\_x(n))$;

        $l\_x \leftarrow lx + 1$;

    end

end

Step 4: Find a threshold to get sub-array spacings

$l\_xm \leftarrow$ mean (unique (n_x))

step 5: Find $xCenters$ corresponding to sub-array spacings

Let $c \leftarrow 1$; Assign $gxc(c) \leftarrow xCenters(1)$;

$gxc(end) \leftarrow xCenters(end)$; $c \leftarrow 2$;

for $j \leftarrow 1$ **to** $length(dist\_x)$ **do**

    if $dist\_x(j) > l\_xm$ **then**

        $dist\_xx(c - 1) \leftarrow dist\_x(j)$;

        $gxc(c) \leftarrow xCenters(j)$;

        $c \leftarrow c + 1$;

    end

end

Step 6: Obtain the average of sub-array spacings

$SubgridDist\_X \leftarrow$ mean$(dist\_xx)$

Step 7: Calculating global grid points

$xG(1) \leftarrow gxc(1) - SubgridDist\_X/2$;

$xG(2 : length(gxc)) \leftarrow gxc(2 : end) + SubgridDist\_X/2$

---

and these should be removed before the global gridding. Three such points will be checked and deleted based on the condition in step 2. Here, $dist\_x$ is the difference between $xCenters$ and $spot\_x$ is the average spot gap between the spots. We have checked the initial 3 point's ($xCenters$) spot gap in $dist\_x$. If the point is less than t_x1 or greater than t_x2, then that is not a valid point. Ideally, the threshold points t_x1 and t_x2 should be $spot\_x$ equivalent. But we can not give the ideal measure in real images. Hence, 20 mm offset to $spot\_x$ is experimentally determined and hence selected for all the images in the four datasets. An example
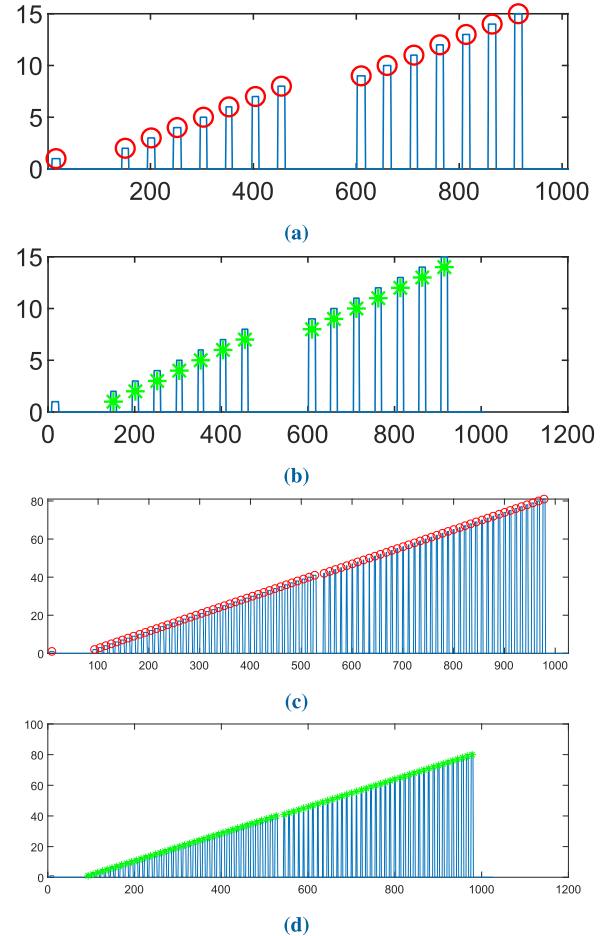

FIGURE 9. Images are Def665 from SIB and OD046 from DeRisi dataset, number of red circled peaks in (a) and (c) is the count of original *xCenters*; but *xCenters*(1) (first point) is a noisy one and that is not considered in refined *xCenters* that are plotted as green starred on the same peaks in (b) and (d).

of plotting noisy and refined $xCenters$ is illustrated in Fig. 9. Then $dist\_x$ is updated with the refined inter spot spacings of $xCenters$. Usually, sub-array spacings are greater than the spot spacings. In order to get correct sub-array spacings, a new threshold l_xm is calculated in steps 3 and 4. Since the noisy points are already removed in step 2, $xCenters(1)$ and $xCenters(end)$ will be the new beginning and endpoints of global $xCenters$ ($gxc(1)$ and $gxc(end)$). Distances or spacings greater than l_xm are counted as sub-array spacings in the loop in step 5 and are stored in an array $dist\_xx$. The remaining global $xCenters$ corresponding to $dist\_xx$ are also taken in the same loop. $SubgridDist\_X$ is calculated as the mean of $dist\_xx$ in step 6. The scale of $SubgridDist\_X$ is added to each global $xCenters$ in step 7 to align sub-arrays effectively within the middle region of the global grid compartments.

In step 7, xG gives the vertical grid lines for sub-arrays. Similarly yG can be obtained for horizontal grid lines using the above algorithm by loading $yCenters$. Region of interest (ROI) around each subarray is a 4 element vector. It contains positions in x and y directions with sizes dx and dy.

TABLE 2. The values of three quality parameters applied to analyze the traits of our experimental datasets are listed. $Niqe_{old}$ and $Niqe_{new}$ are the niqe values of source image and enhanced image 'ze'; Similarly $Cont_{old}$ and $Cont_{new}$ are the contrast degree values of source image and image 'ze'; Note: $RMS_C * 100$ values are shown here. A source image is the grayscale version of the RGB image obtained from the original red channel and green channel images. The blue channel is kept as zero.

Dataset	Images	mse_g	mse_r	$Niqe_{old}$	$Niqe_{new}$	$Cont_{old}$	$Cont_{new}$	$Cont_{diff}$	Best Rot Angle_x	Best Rot Angle_y
SIB	661	17.20	5.52	5.17	10.08	4.41	13.47	9.06	0.3	0
	662	7.38	10.71	5.75	23.08	4.53	13.56	9.03	1.1	-0.8
	663	22.77	10.38	6.4	16.51	4.93	13.97	9.03	0.6	-0.2
	664	12.53	18.59	4.47	16.51	3.06	15.1	12.03	0.1	-0.7
	665	22.10	316.01	5.64	21.86	4.59	16.91	12.32	0.2	-0.6
	666	8.63	11.64	3.7	19.6	4.47	13.68	9.21	1.4	-1.1
	667	5.44	5.75	4.26	18.32	4.02	13.66	9.64	0.7	-0.2
Derisi	OD370	33.20	29.23	6.13	17.75	5.17	26.71	21.54	0	0.1
	OD046	48.33	53.02	4.69	12.49	5.55	25.79	20.24	0.2	-0.2
	OD014	33.51	44.03	5.13	14.01	6.02	25.51	19.49	0.1	-0.1
	OD180	25.26	29.61	5.25	9.93	4.84	26.43	21.58	-0.1	0.1
	OD690	42.39	45.94	5.25	11.46	4.67	25.21	20.53	0.1	-0.1
	OD080	31.98	43.55	4.94	11.94	5.28	25.60	20.31	-0.3	0.3
	OD730	30.51	31.56	5.13	13.09	2.76	23.56	20.79	0.1	-0.1
GEO	GSM16391	5.25	2.05	2.63	18.29	1.81	12.01	10.19	-0.3	-0.2
	GSM15898	5.17	6.33	2.78	15.89	2.7	11.81	9.11	-0.4	-0.1
	GSM16101	7.16	2.58	1.68	18.51	1.93	13.55	11.61	-0.3	-0.1
	GSM16389	6.91	2.36	1.88	18.3	1.88	12.18	10.30	-0.2	-0.2
SMD	TB1_037,_038	1.52	8.39	4.27	12.07	0.15	17.53	17.37	-0.1	0.1
	TB6_99,_100	0.25	0.08	5.07	8.55	1.04	16.47	15.43	0.1	0.2
	TB3_65,_67	0.31	21.09	4.48	8.99	0.65	14.05	13.39	0.0	0.3
	TB3_89,_90	8.19	3.75	4.35	11.81	0.17	14.61	14.43	0.0	0.2
	TB3_96,_94	0.08	14.10	5.64	9.37	0.41	15.03	14.61	0.0	0.2
	TB3_95,_97	0.15	6.13	3.85	8.5	0.59	16.63	16.03	0.0	0.3
	TB3_101,_102	1.65	9.84	5.17	11.23	0.17	14.48	14.31	-0.2	0.3
	Microti_5_95,_3_96	138.20	106.19	2.71	11.57	4.75	20.14	15.38	0.7	0.0
	2001-_0008.1, 8_0009.1	35.07	35.85	3.56	11.16	1.33	19.14	17.81	0.0	0.1
	2001-_0010, 0010.1	1.18	3.41	4.39	10.48	0.38	20.65	20.26	0.0	0.2

E. PROPOSED LOCAL GRIDDING

Major steps of the Local gridding algorithm are listed in Algorithm 3. Steps 1 and 2 do the procedure as in the same steps of the global gridding algorithm. Load the $xCenters$ from subgrid's ROI's and find the optimal spot spacing $spot\_x$ as seen in step 1. In step 2, threshold values $tloc\_x1$ (lower limit) and $tloc\_x2$ (upper limit) are obtained from the $spot\_x$ to obtain the correct grid lines. Here, 5 mm offset to $spot\_x$ is experimentally determined and hence selected for all the images in the four datasets. All adjacent $xCenters$ are considered correct points if their separation distance is within the defined lower and upper limits. First and endpoints are checked based on the conditions in step 2 to identify the noisy $xCenters$. $xCenters(1)$ is taken into the new variable $locx(1)$ to continue the remaining steps. Step 3 does the following

cases to regulate the grid line for proper gridding of all of the spots. Case 1: if $dist\_x(i)$ is within the limits of $tloc\_x1$ and $tloc\_x2$, then the next $xCenters(i+1)$ is the correct point and will be considered. Case 2: if $dist\_x(i)$ is lesser than the lower limit $tloc\_x1$ and $dist\_x(i+1)$ is within the limits of $tloc\_x1$ and $tloc\_x2$, then current point and next point should be considered. This will be executed in an algorithm when the variable $count \leftarrow 1$. Case 3: if both $dist\_x(i)$ and $dist\_x(i+1)$ are lesser than the $tloc\_x1$, then $xCenters(i+1)$ is assumed as a noisy point and i is incremented by one to take the next $xCenters(i+1)$. This happens when $count \leftarrow 2$ in the algorithm. Case 4: if $dist\_x(i)$ is greater than $tloc\_x2$, obviously one or more $xCenters$ are missed between adjacent points. We will calculate the number of points missed by the equation $num \leftarrow round(dist\_x(i)/spot\_x)$, and correspond-

ing points will be introduced in the loop for the equation $locx(j) \leftarrow xCenters(i+1) - check * spot\_x$. Hence, the proposed technique is appropriate to microarray images gridding with different levels of hybridization, and in the absence of spots. The obtained $locx$ are re-assigned back to $xCenters$. Step 4 finds the distances to be added to that $xCenters$ to obtain the grid lines. The grid lines are computed in step 5 in such a way that the points are residing within the middle region of each grid compartment.

In step 5, $xGrid$ gives the vertical grid lines for the spots. Similarly, $yGrid$ can be obtained for horizontal grid lines using the above algorithm by loading $yCenters$.

V. RESULTS AND DISCUSSION

Three types of noises namely Salt and pepper noise, Poisson noise, and Gaussian noise are commonly found in the microarray images. Sparsely occurring white pixels in dark areas and black pixels in white areas can be eliminated by a median filter. This noise is mainly because of high transients in the signal that results in ADC errors during the conversion from an analog signal to a digitized image. Poisson noise is associated with the variation of the number of photons sensed at a given exposure level. The exposure level is defined by green and red wavelengths of laser medium to the corresponding dyes of normal and abnormal DNA samples. The photomultiplier tubes in the image acquisition system multiply the number of electrons of sensed photons from the DNA chip which creates high amplification noise (considered as Gaussian noise) in order to enhance the low light conditions. The preprocessing algorithm proposed in Algorithm 1 gives 'z' which is top-hat filtered, and followed by median filtered to give a denoised and enhanced image 'z_e'.

This 'z_e' is passed to the Global gridding stage. Global gridding lines vertical (V), horizontal (H) are calculated in the horizontal and in the vertical direction according to Algorithm 2. The measurements obtained from the Global gridding algorithm is provided in Table 3. The required, obtained correct, obtained false and obtained total V, H lines for each image in the four databases are computed. Also, average inter-subgrid distances in x and y direction are added in the table. The accuracy of global gridding lines is computed by (3).

$$Accuracy = \frac{sum(obtained\ correct)}{sum(obtained\ total)} \quad (3)$$

The comparison of our global gridding results with previous works which used the same datasets are displayed in Table 4. The authors in [27] proposed a fully automatic gridding algorithm and had worked on the Derisi dataset. However, they only gave the local gridding results of that dataset.

SMD [1] used in our work is different from SMD [2] used in the literature. Both are SMD microarray databases with different numbers of sub-arrays, spots, and other features.

Algorithm 3: Proposed Local Gridding Algorithm

Result: Refined Local vertical gridding points

Step 1: Load $xCenters$ from subgrid's ROI's of z_e;

Let $dist\_x \leftarrow$ differences of $xCenters$;

$spot\_x \leftarrow$ median of $dist\_x$

Step 2: Refining $xCenters$ to check noisy points;

Let $tloc\_x1 \leftarrow (spot\_x - spot\_x/5)$ and
$tloc\_x2 \leftarrow (spot\_x + spot\_x/5)$;
$h \leftarrow 1; \ j \leftarrow 2; \ count \leftarrow 0$

Checking first point:

if $xCenters(h) < spot\_x/2$ **then**
  $locx(h) \leftarrow xCenters(2)$;
  $h \leftarrow h + 1$;
end

Checking end point:

if $dist\_x(end) > tloc\_x2$ **then**
  delete $xCenters(end)$;
end

$dist\_x \leftarrow$ differences of new $xCenters$

Step 3: Regulating $xCenters$ for correct gridding points;

for $i \leftarrow h$ **to** $length(dist\_x)$ **do**
  if $dist\_x(i) >= tloc\_x1$ and $dist\_x(i) <= tloc\_x2$
  then
    if $count = 1$ **then**
      $locx(j) \leftarrow xCenters(i)$;
      $j \leftarrow j + 1$;
      $count \leftarrow 0$;
    end
    $locx(j) \leftarrow xCenters(i+1)$;
    $j \leftarrow j + 1$;
  else if $dist\_x(i) < tloc\_x1$ **then**
    $count \leftarrow count + 1$;
  else if $dist\_x(i) > tloc\_x2$ **then**
    $num \leftarrow round(dist\_x(i)/spot\_x)$;
    $check \leftarrow num - 1$;
  for $n \leftarrow 1$ **to** num **do**
    $locx(j) \leftarrow xCenters(i+1) - check * spot\_x$;
    $j \leftarrow j + 1$;
    $check \leftarrow check - 1$;
  end
  end
  if $count = 2$ **then**
    $locx(j) \leftarrow xCenters(i+1)$;
    $j \leftarrow j + 1$;
    $count \leftarrow 0$;
  end
end

Clear $xCenters$; Assign updated $xCenters \leftarrow locx$

Step 4: Find $gap \leftarrow differences(xCenters)/2$

Step 5: Calculating local gridding points;

$xGrid\ (1) \leftarrow xCenters(1) - gap(1)$;

$xGrid\ (2 : length(xCenters) + 1) \leftarrow$
$xCenters(1 : end) + gap[1 : end\ end]$

TABLE 3. Measurements obtained from the proposed Global gridding algorithm. Note: Avg Subgrid dist_X is the average value of obtained sub-array/subgrid distances in the horizontal profile. Similarly, Avg Subgrid dist_Y is the average value of obtained sub-array/subgrid distances in the vertical profile. Accuracy = sum (Obtained correct)/ sum (Obtained total).

Dataset	Images	Required V, H	Obtained correct V, H	Obtained false V, H	Obtained total V, H	Avg Subgrid dist_X (mm)	Avg Subgrid dist_Y (mm)
SIB	661	3,3	3,3	0,0	3,3	155.5	147.5
	662	3,3	3,3	0,0	3,3	155	148
	663	3,3	3,3	0,0	3,3	154.5	148.5
Tough	**664**	**3,3**	**3,2**	**0,1**	**3,3**	154	149
Tough	**665**	**3,3**	**3,3**	**0,0**	**3,3**	155	148
	666	3,3	3,3	0,0	3,3	155.5	147
	667	3,3	3,3	0,0	3,3	154	148
	Total	21+21=42	21+20=41		21+21=42	**Accuracy**	**41/42=0.98**
DeRisi	OD180	3,3	3,3	0,0	3,3	18.5	15.5
Tough	OD370	3,3	3,3	0,0	3,3	18.5	15
Tough	OD046	3,3	3,3	0,0	3,3	18.5	14.5
	OD014	3,3	3,3	0,0	3,3	18.5	14.5
	OD690	3,3	3,3	0,0	3,3	19	15
	OD080	3,3	3,3	0,0	3,3	18.5	15.5
	OD730	3,3	3,3	0,0	3,3	18.5	15
	Total	21+21=42	21+21=42		21+21=42	**Accuracy**	**42/42=1**
GEO	GSM16391	5,13	5,13	0,0	5,13	190	225.5
Tough	GSM15898	5,13	5,13	0,0	5,13	191.16	238.04
	GSM16101	5,13	5,13	0,0	5,13	190.16	220.5
	GSM16389	5,13	5,13	0,0	5,13	190.16	233.4
	Total	20+52=72	20+52=72		20+52=72	**Accuracy**	**72/72=1**
SMD	TB6_99,_100	5,5	5,5	0,0	5,5	66.67	100.34
Tough	TB1_37,_38	5,5	5,5	3,2	8,7	58	121.3
	TB3_65,_67	5,5	5,4	0,1	5,5	66.83	162.83
	TB3_89,_90	5,5	5,5	0,1	5,6	67.16	107.25
	TB3_96,_94	5,5	5,5	0,1	5,6	66.50	135.62
	TB3_95,_97	5,5	5,4	0,1	5,5	67	162
	TB3_101,_102	5,5	5,5	0,1	5,6	66.67	124.67
Tough	Mic_95,_96	5,5	4,5	3,1	7,6	69.70	133.37
Tough	8.1,8_0009.1	5,5	5,5	1,0	6,5	62.12	115.50
	0010,0010.1	5,5	5,5	0,0	5,5	67.16	145.67
	Total	50+50=100	49+48=97		56+56=112	**Accuracy**	**97/112=0.87**

Since SMD [2] is not available for our work, we neglected the comparison of SMD's results in local gridding results and discussion. The resultant figures after the global gridding are exhibited in Fig. 10. Global gridding results are all plotted on corresponding enhanced images 'z_e's. A source image is the grayscale version of the RGB image obtained from the original red channel and green channel images. ROI around each sub-array is a 4 element vector which can be calculated using horizontal and vertical grid lines attained from the global gridding stage.

ROIs of sub-arrays are automatically passed into the local gridding stage. Local gridding lines are computed according to Algorithm 3. Measurements from the local gridding algorithm are listed in Table 5. The total number of spots (T) is the overall number of spots in the entire image. True Positive (TP) is the number of spots that are perfectly gridded. False Positive (FP) is the number of background regions that are wrongly gridded as spot blocks. The number of spots missed and considered as the background is counted in False Negative (FN). In our work, 0% is the FP rate. 21 spots
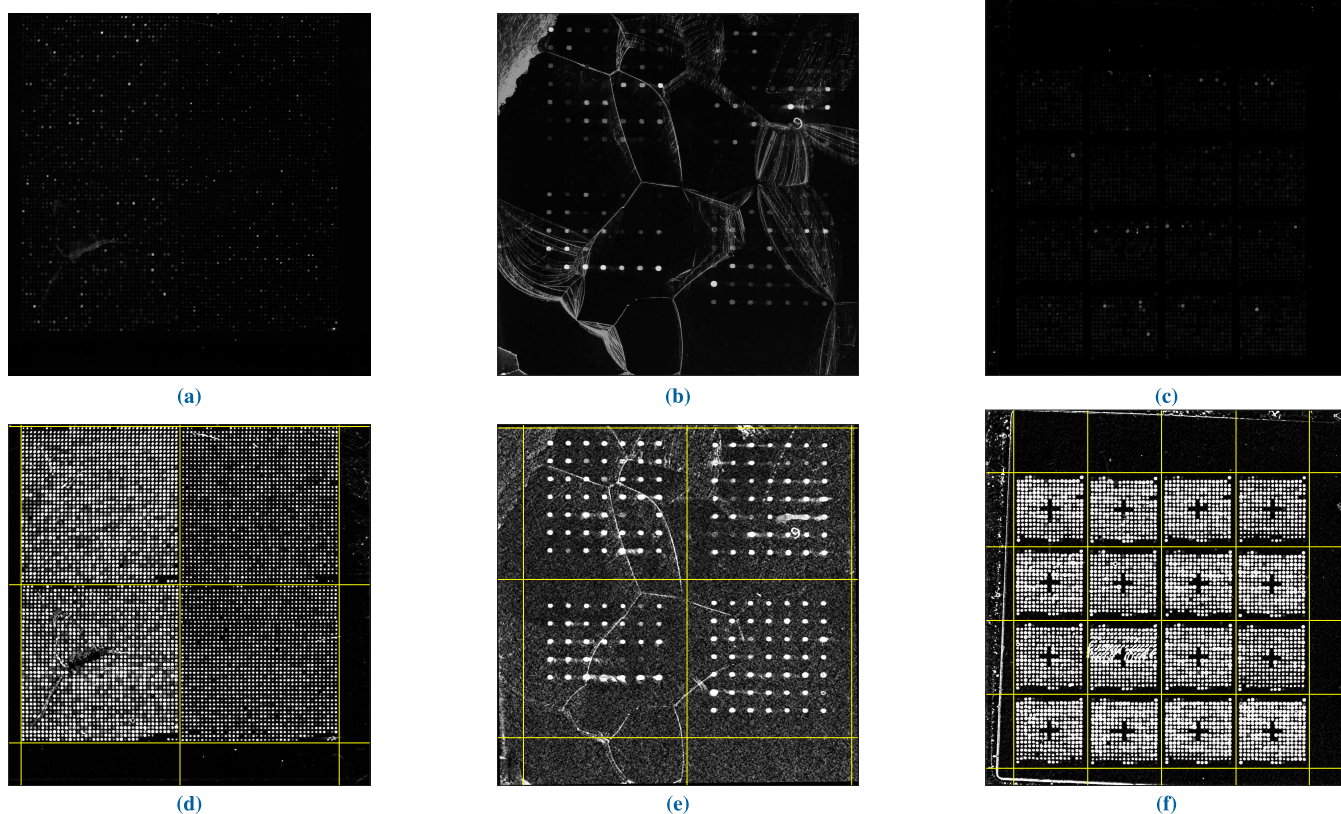
FIGURE 10. (a) and (d) Images OD730 (DeRisi dataset), (b) and (e) Def665 (SIB dataset), (c) and (f) 2001-_0010 (SMD dataset), (a), (b), and (c) are the source images and (d), (e), and (f) are their globally gridded enhanced images.

TABLE 4. Comparison of accuracy in % (accuracy*100) of the proposed Global gridding algorithm with competitive works.

Dataset	Rueda [17]	Mary Monir [27]	Our Method
Year	2011	2019	2019
SIB	-	89.29%	98%
Derisi	-	-	100%
Geo	100%	100%	100%
SMD [1]	-	-	87%
SMD [2]	100%	-	-

each from sub-array 1 and 3 of Def 664 are missed and fallen into the FN group due to the false top horizontal global gridding line. This is depicted in Fig. 11. Image Def664 is globally gridded 98% and locally gridded 76.93% with our method. Image Def665 is globally and locally gridded with 100% and is displayed in Fig. 12. Once the gridding is perfectly aligned, the grid lines can be superimposed on source images or enhanced images. Fig. 13 and Fig. 14 shows the local gridding results of images from Derisi and GEO datasets. The number of FN is more observed in the GEO dataset. A large number of spots in the top as well as in the bottom lines of sub-arrays in the GEO dataset are missed due to the low value of horizontal mean intensity in those lines. A comparison in terms of accuracy with most competitive works is shown in Table 6. This is also illustrated in bar
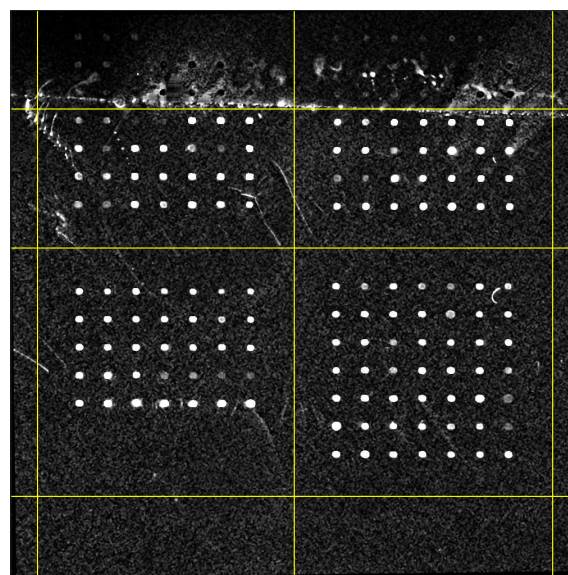

FIGURE 11. False top horizontal gridding line attained in the global gridding algorithm in Image Def664 (SIB dataset).

diagrams in Fig. 15. The authors of [10], [17] and [27] did gridding on entire images and authors of [14], [19] and [20] worked on cropped subgrids.

In our work, the accuracy of perfectly gridded spots is only taken into the account. The accuracy of local gridding lines
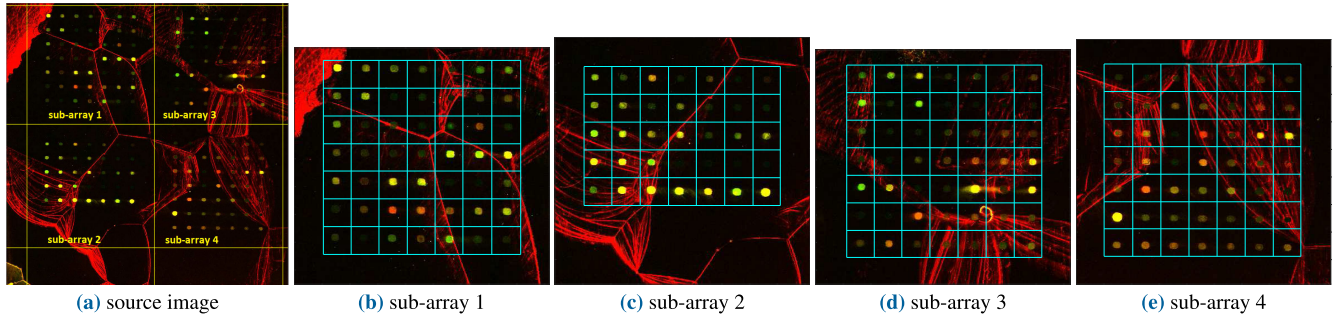
| (a) source image | (b) sub-array 1 | (c) sub-array 2 | (d) sub-array 3 | (e) sub-array 4 |

FIGURE 12. Image Def665 from SIB dataset. (Note: obtained grid lines are superimposed on source image in rgb domain).

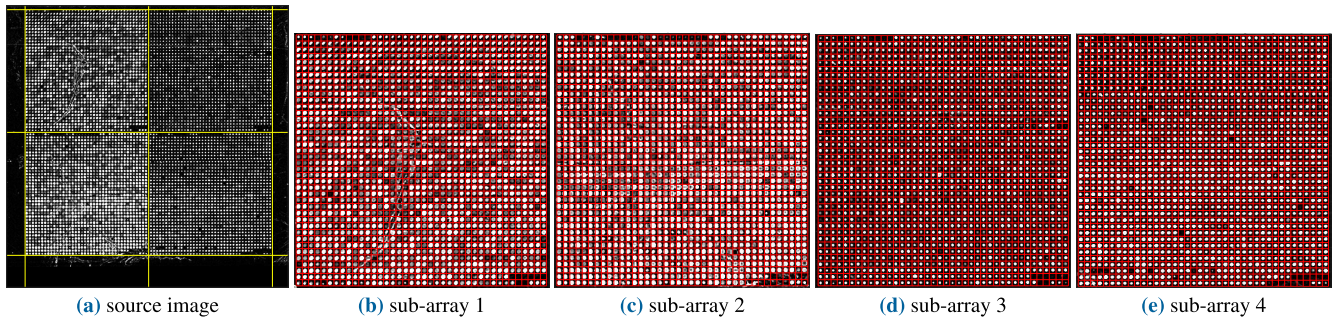| (a) source image | (b) sub-array 1 | (c) sub-array 2 | (d) sub-array 3 | (e) sub-array 4 |

FIGURE 13. Image OD046 from DeRisi dataset. (Note: grid lines are plotted on enhanced subarrays).

is defined by (4).

$$Accuracy = \frac{TP}{TP + FP + FN} \qquad (4)$$

Accuracy for each image is given independently in Table 5. In the SIB dataset, except Def 664, other images got 100% accuracy. Def664 has 76.93% (FN value = 42). DeRisi dataset achieved 100% accuracy for all the images. GEO dataset attained 97.85% for GSM16101, 96.1% for GSM15898, 98.45% for GSM16389 and 99.2% for GSM16391 accuracy respectively. The comparison of our local gridding results with the previous works is exhibited in Figs. 16, 17, and 18. The local gridding algorithm gives an overall accuracy of 96.7%, 100% and 97.9% for SIB, DeRisi and GEO datasets respectively.

The proposed entire algorithm is appropriate to microarray images gridding with different levels of hybridization and in the absence of spots. All the experimental databases have different levels of spot hybridization. Every spot intensity is proportional to that particular spot's degree of hybridization of green-colored genes (control genes) and red-colored genes (test genes). The absence of a spot or missing spot is due to the lack of that particular gene being hybridized. This algorithm works effectively in a single row/column in a sub-array for the case of a large number of missing spots or non-hybridized spots. GEO databases have a lot of sub-arrays, with a lot of missed spots. Even with that database, we got a better gridding accuracy than the existing methods.
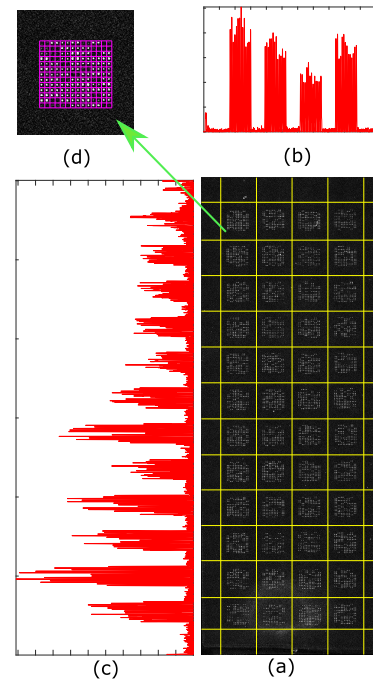

FIGURE 14. Image GSM16389 from Geo dataset, (a) globally gridded on enhanced image, (b) enhanced horizontal profile, (c) enhanced vertical profile, and (d) locally gridded on sub-array 1.

A. COMPUTATIONAL EFFICIENCY PERFORMANCE

Microarray image-oriented analysis is usually not performed in real-time. The importance lies more on how easily and correctly we can interpret the intensity values of the microar-

TABLE 5. Measurements obtained from the proposed Local gridding algorithm. Note: Avg Spot dist_X is the average value of obtained inter spot distances in the horizontal profiles of all subgrids. Similarly, Avg Spot dist_Y is the average value of obtained inter spot distances in the vertical profiles of all subgrids. *Accuracy* $= \frac{TP}{TP+FP+FN} * 100$.

| Dataset/ Main challenges	Images	Avg Spot dist_X(mm)	Avg Spot dist_Y(mm)	Total spots (T)	True Positive (TP)	False Positive (FP)	False Negative (FN)	Accuracy in %
SIB	Def661	50.5	50	182	182	0	0	100%
Block Noise in subgrid 4	Def662	51	50	182	182	0	0	100%
Noises in b/n column of spots of subgrid 3	Def663	51	49.75	182	182	0	0	100%
Low quality spots in top 3 rows	Def664	51	49.5	182	140	0	42	76.93%
Large noise due to experimental variations	Def665	50.75	50	182	182	0	0	100%
Big Tilt [1.4,-1.1]	Def666	51	50	182	182	0	0	100%
	Def667	51	50	182	182	0	0	100%
		Overall Accuracy	**96.70%**					
Derisi	OD180	11	11	6400	6400	0	0	100%
Large no of missing spots in row 1	OD370	11	11	6400	6400	0	0	100%
Noise in subgrid 1, missing spots in subgrid 2	OD046	11	11	6400	6400	0	0	100%
	OD014	11	11	6400	6400	0	0	100%
Low quality spots in subgrid 2	OD690	11	11	6400	6400	0	0	100%
Large no of missing spots in subgrid 4	OD080	11	11	6400	6400	0	0	100%
Low quality spots in subgrid 2	OD730	11	11	6400	6400	0	0	100%
		Overall Accuracy	**100%**					
GEO	GSM16101	20	19	8736	8548	0	188	97.85%
Large no of missing spots	GSM15898	20	20	8736	8394	0	342	96.1%
	GSM16389	20	19.5	8736	8600	0	136	98.45%
	GSM16391	20	18.75	8736	8659	0	77	99.2%
		Overall Accuracy	**97.9%**					
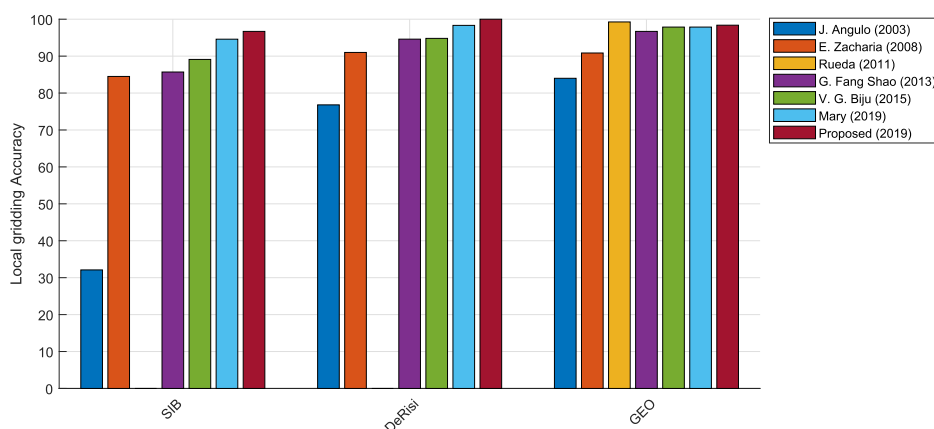
FIGURE 15. Comparison in bar diagrams of local gridding accuracy.

ray spots given the noises and artifacts that occur from experimental variations and inappropriate imaging conditions. After acquiring or storing the data, the image analysis can be done offline. We have computed and executed our algorithms in Matlab R2018b and ran on an Intel-based workstation with clock speed at 3.30 GHz. The operating system is Windows 10 Pro edition with 16 GB installed RAM. The average processing time of the entire image (complete spots in a red channel image and in a green channel image) of the proposed fully automated technique is calculated as 3.5 seconds for the SIB dataset, 6 seconds for the DeRisi dataset, 30 seconds for the GEO dataset and 12 seconds for the SMD
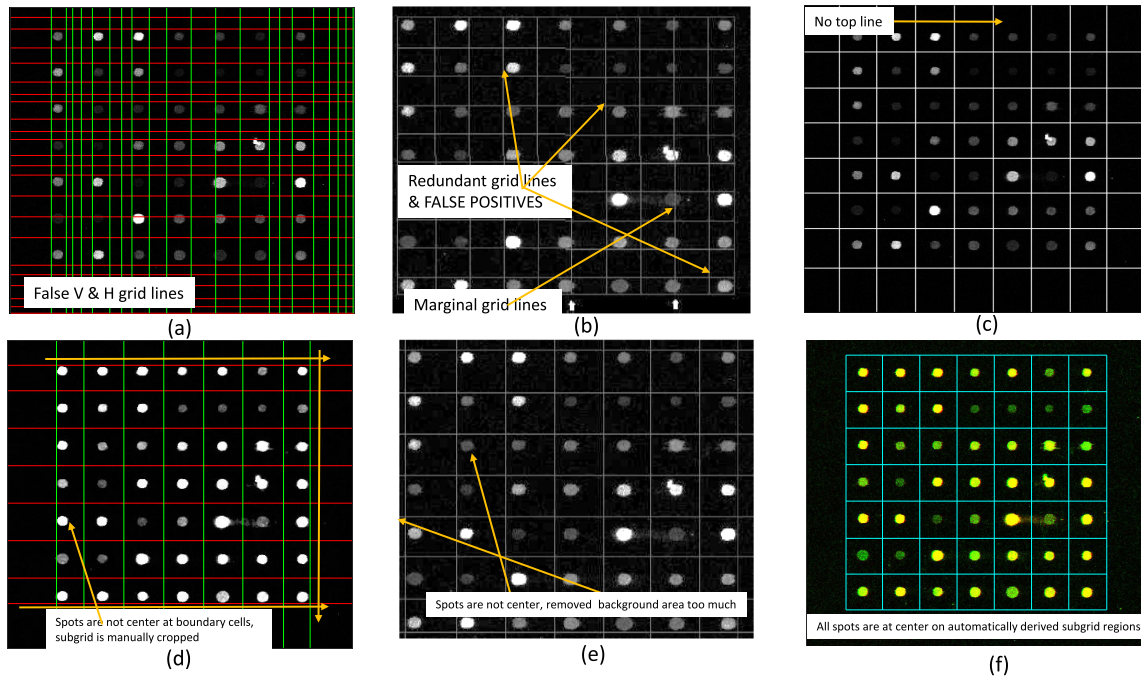
FIGURE 16. Comparison on image Def661 subgrid 1, (a) J. Angulo (Morphology method), (b) E. Zacharia (Genetic Algorithm), (c) Rueda (OMTG method), (d) G. Shao (Modified Otsu method), (e) V. G. Biju (Gridline refinement method), and (f) Proposed method.
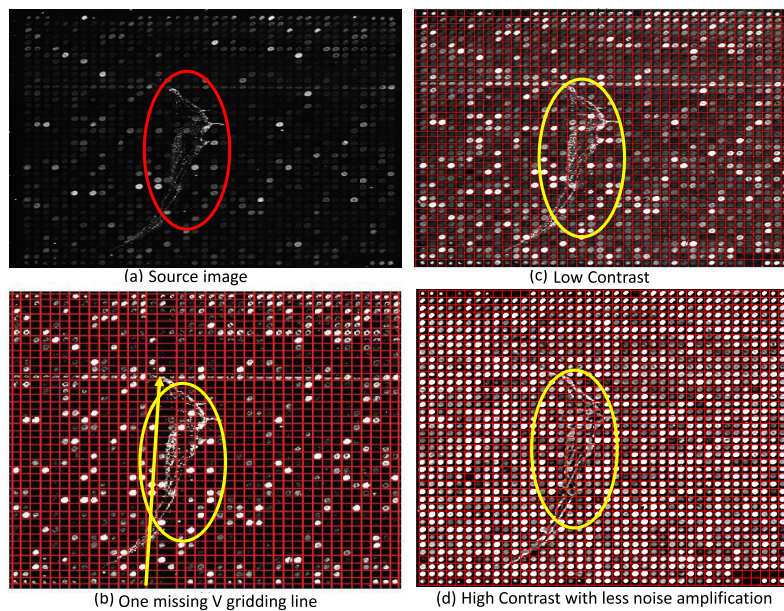

FIGURE 17. Comparison on image OD046 subgrid 1, (a) Source image, (b) Shao method, (c) Monir method, and (d) Proposed method. Yellow circles show the noise effect on their enhanced images.

TABLE 6. Comparison of the accuracy of the proposed Local gridding algorithm with the competitive works. Accuracy is also compared with the works applied on cropped subgrids and are displayed in the table.

Dataset	Jesus Angulo [10]	Eleni Zacharia [14]	Rueda [17]	Guifang Shao [19]	V. G. Biju [20]	Mary Monir [27]	Our Method
Year	2003	2008	2011	2013	2015	2019	2019
SIB	32.1%	84.5%	-	85.7%	89.11%	94.59%	96.7%
Derisi	76.8%	91%	-	94.6%	94.81%	98.33%	100%
Geo SMD [2]	84% 92%	90.85% 92.6%	99.26% 98.06%	96.7% 98.9%	97.87% 96.51%	97.87% -	97.9% -

(a) Noise amplified along with spots   (b) False grid cells at boundary   (c) Less noise effect on enhanced image
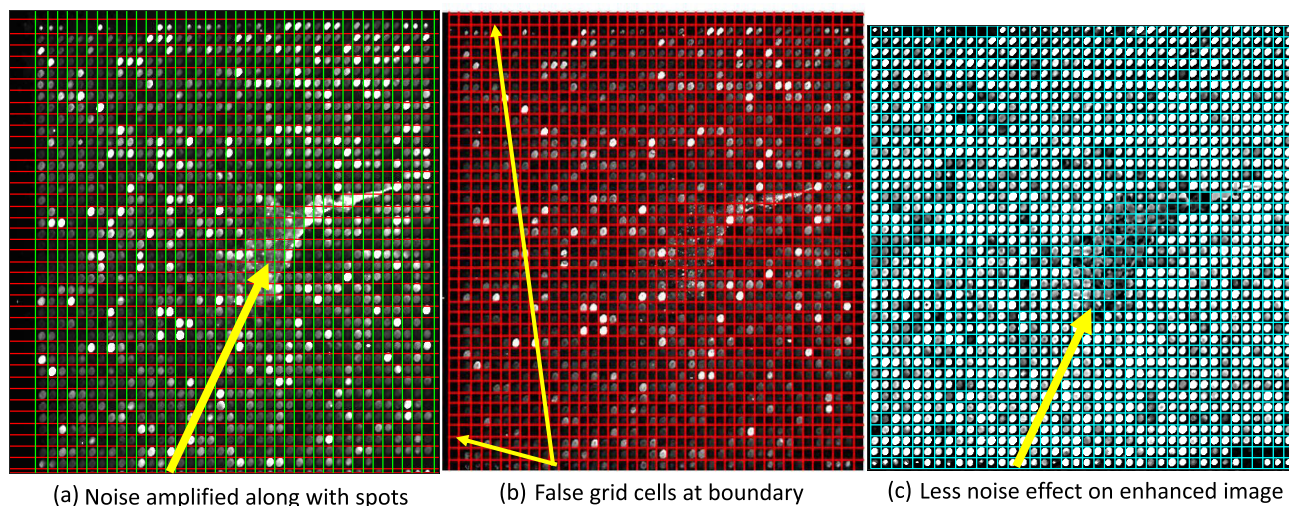
FIGURE 18. Comparison on image OD080 subgrid 1, (a) Shao method, (b) Monir method, and (c) Proposed method.

TABLE 7. Average Processing speed for the proposed technique including three algorithms for the experimental datasets.

Dataset	Average Processing speed in seconds(s)
SIB	3.4 s
Derisi	3.5 s
Geo	30 s
SMD [1]	12 s

dataset respectively. That is shown in Table 7. Matlab tic-toc commands are used to find the elapsed time for the four experimental datasets to implement the proposed technique in full. Quality in computational efficiency differs with various datasets, software and device configurations for microarray images. Therefore, the processing speed is different. Due to the non-availability of the complete algorithm/codes of other state-of-the-art methods, we could not compare the processing time of our method with those methods. Moreover, in most of the available methods, a fully automated gridding algorithm on an entire image is not developed (only the average processing speed for a single spot in a sub-array or for a single sub-array is calculated).

VI. CONCLUSION

In this work, a composite microarray image is formed by suitably stacking red and green channels to make the whole image analysis algorithm computationally efficient. The proposed preprocessing method, which combines the advantages of median filtering in the 3D domain and contrast enhancement by dehazing, increases the contrast of the entire image with reduced noise effect. The resultant enhanced image after top-hat filtering and median filtering attains a better contrast with the reduced effect of noise amplification. Many of the fully automated works available in the literature are based on the cropped subgrids of images. The proposed fully automatic

algorithm for global gridding combined with local gridding uses the structural information as features to obtain the correct gridding lines. Gridline management is performed in a row/column of an image to find the missing spots or the non-hybridized spots. Hence, the proposed technique is appropriate to microarray images gridding with different levels of hybridization, and in the absence of spots. The traits of a microarray image are evaluated using three parameters namely Mean square error, Naturalness quality image evaluator and degree of contrast. No user intervention or preset parameters are introduced in the proposed method. Tough images in SIB, DeRisi, GEO, and SMD are gridded with better accuracy than the previous works in the literature. We obtained a 0% false positive (FP) rate for the experimental datasets using the proposed method. The highest global gridding accuracy of the SIB dataset reported so far is 89. 29% [27] whereas proposed method gives 98% accuracy. Also, we achieved 2.11%, 1.67% and .03% improvement in local gridding accuracy for SIB, Derisi and GEO datasets than the aforementioned latest work.

REFERENCES

[1] 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, Mar. 2002.

[2] S. M. Joseph and P. S. Sathidevi, "CDNA microarray image enhancement for effective gridding of spots," in *Proc. IEEE Region 10 Conf. (TENCON)*, Oct. 2019, pp. 326–331.

[3] Genetic Science Learning Center. (Oct. 23, 2018). *DNA Microarray, Learn. Genetics*. Accessed: Feb. 21, 2019. [Online]. Available: https://learn.genetics.utah.edu/content/labs/microarray/

[4] M. B. Eisen, *ScanAlyze User Manual*. Stanford, CA, USA: Stanford Univ., 1999.

[5] *GenePix 4000A User's Guide*, Axon Instrum., Union City, CA, USA, 1999.

[6] P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush, "A concise guide to cDNA microarray analysis," *BioTechniques*, vol. 29, no. 3, pp. 548–562, Sep. 2000.

[7] *6.1 User Manual*, I. ImaGene, 2008.

[8] J. Buhler, T. Ideker, and D. Haynor, "Dapple: Improved techniques for finding spots on dna microarrays," CSE, Univ. Washington, Seattle, WA, USA, Tech. Rep. 112, Aug. 2000, pp. 5–8.

[9] M. Katzer, F. Kummert, and G. Sagerer, "A Markov random field model of microarray gridding," in *Proc. ACM Symp. Appl. Comput. (SAC)*, 2003, pp. 72–77.

[10] J. Angulo and J. Serra, "Automatic analysis of DNA microarray images using mathematical morphology," *Bioinformatics*, vol. 19, no. 5, pp. 553–562, Mar. 2003.

[11] P. Bajcsy, "Gridline: Automatic grid alignment in DNA microarray scans," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 15–25, Jan. 2004.

[12] L. Rueda and V. Vidyadharan, "A hill-climbing approach for automatic gridding of cDNA microarray images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 3, no. 1, pp. 72–83, Jan. 2006.

[13] M. Ceccarelli and G. Antoniol, "A deformable grid-matching approach for microarray images," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3178–3188, Oct. 2006.

[14] E. Zacharia and D. Maroulis, "An original genetic approach to the fully automatic gridding of microarray images," *IEEE Trans. Med. Imag.*, vol. 27, no. 6, pp. 805–813, Jun. 2008.

[15] S. Katsigiannis, E. Zacharia, and D. Maroulis, "MIGS-GPU: Microarray image gridding and segmentation on the GPU," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 3, pp. 867–874, May 2017.

[16] D. Bariamis, D. K. Iakovidis, and D. Maroulis, "M3G: Maximum margin microarray gridding," *BMC Bioinf.*, vol. 11, no. 1, p. 49, 2010.

[17] L. Rueda and I. Rezaeian, "A fully automatic gridding method for cDNA microarray images," *BMC Bioinf.*, vol. 12, no. 1, p. 113, Apr. 2011.

[18] J. Deepa and T. Thomas, "A new gridding technique for high density microarray images using intensity projection profile of best sub image," *Comput. Eng. Intell. Syst.*, vol. 4, no. 1, pp. 7–18, 2013.

[19] G.-F. Shao, F. Yang, Q. Zhang, Q.-F. Zhou, and L.-K. Luo, "Using the maximum between-class variance for automatic gridding of cDNA microarray images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, no. 1, pp. 181–192, Jan. 2013.

[20] V. G. Biju, and P. Mythili, "Microarray image gridding using grid line refinement technique," *ICTACT J. Image Video Process.*, vol. 5, no. 4, pp. 1010–1016, May 2015.

[21] N. Giannakeas, F. Kalatzis, M. G. Tsipouras, and D. I. Fotiadis, "A generalized methodology for the gridding of microarray images with rectangular or hexagonal grid," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 719–728, Jul. 2015.

[22] N. Giannakeas, F. Kalatzis, M. G. Tsipouras, and D. I. Fotiadis, "Spot addressing for microarray images structured in hexagonal grids," *Comput. Methods Programs Biomed.*, vol. 106, no. 1, pp. 1–13, Apr. 2012.

[23] N. Dehghan Khalilabad and H. Hassanpour, "Employing image processing techniques for cancer detection using microarray images," *Comput. Biol. Med.*, vol. 81, pp. 139–147, Feb. 2017.

[24] H. Saberkari, M. Shamsi, and H. B. Ghavifekr, "A shape-independent algorithm for fully-automated gridding of cDNA microarray images," *Comput. Electr. Eng.*, vol. 62, pp. 135–150, Aug. 2017.

[25] S. Karthik and S. Manjunath, "An automated and efficient approach for spot identification of microarray images using X-covariance," in *Proc. Int. Conf. Cognition Recognit.* Singapore: Springer, pp. 273–282, 2018.

[26] Z. Gan, N. Zeng, F. Zou, J. Chen, M. Du, L. Liao, H. Li, and Y. Zhang, "Multilevel segmentation optimized by physical information for gridding of microarray images," *IEEE Access*, vol. 7, pp. 32146–32153, 2019.

[27] M. M. Saeid, Z. B. Nossair, and M. A. Saleh, "A fully automated spot detection approach for cDNA microarray images using adaptive thresholds and multi-resolution analysis," *IEEE Access*, vol. 7, pp. 80380–80389, 2019.

[28] R. Bierman, N. Maniyar, C. Parsons, and R. Singh, "MACE: Lossless compression and analysis of microarray images," in *Proc. ACM Symp. Appl. Comput.*, 2006, pp. 167–172.

[29] *SIB Dataset*. Accessed: Aug. 10, 2017. [Online]. Available: http://www.isrec.isb-sib.ch/

[30] *DeRisi Dataset*. Accessed: Mar. 15, 2019. [Online]. Available: http://www.bio.davidson.edu/projects/magic/magic.html

[31] *GEO Dataset*. Accessed: Mar. 15, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/geo/

[32] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[33] X. Dong, Y. A. Pang, G. Wang, W. Li, Y. Gao, and S. Yang, "Fast efficient algorithm for enhancement of low lighting video," *J. Inf. Comput. Sci.*, vol. 7, no. 10, pp. 2021–2030, 2010.

[34] R. Bemis. (2020). *DNA MicroArray Image Processing Case Study*. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/2573-dna-microarray-image-processing-case-study

[35] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

STEFFY MARIA JOSEPH (Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the Vimal Jyothi Engineering College, Chemperi, Kannur University, in 2011, and the M.Tech. degree in signal processing from Government Engineering College, Kozhikode, Calicut University, in 2015. She is currently pursuing the Ph.D. degree with the National Institute of Technology, Calicut, India. Her research interests include image processing, microarray image analysis, machine learning, and signal compression.

P. S. SATHIDEVI (Member, IEEE) received the B.Tech. and Ph.D. degrees from R. E. C, Calicut, India, and the M.Tech. degree from CUSAT, Kochi, India. She is currently working as a Professor with the Department of Electronics and Communication Engineering, National Institute of Technology, Calicut. Her research interests include speech and audio processing, biomedical signal processing, signal compression, wavelets, and cryptography.

• • •