Microarray images: Defect Analysis

Chinmay Athavale[1]
[1] *M.Tech, Dept. of Electronics engineering, GHRC, Nagpur*
E-mail: athavale_chinmay.ghrcemtechetrx@raisoni.net

Shailesh Bhalerao[2]
[2] *Prof. Dept. of Electronics engineering, GHRC, Nagpur*
E-mail: shailesh.bhalerao@raisoni.net

Abstract—Analysis of complementary-DNA (cDNA) is very useful in diagnosis of genetic traits and viral diseases. Microarray facilitates that using image processing to provide accurate, quick and efficient analysis. Since microarray images are complex, these consist several defects that have to be corrected using several algorithms. In this research work, an innovative model for defect analysis has been proposed to characterise the defects to provide better facility for their correction. Further, defects are classified by their characteristics and cause behind them as well as impact these defects are having on microarray image analysis. Along with above, concept of ideal image is also proposed by defining new parameters that can lead to accurate and efficient analysis.

Keywords— *Defect analysis, Microarray images, Image processing, defect classificatio, Image defects.*

I. INTRODUCTION

Genetic analysis is a key technology in medical diagnosis sector because of its potential to detect viral diseases and genetic defects. Gene analysis often comprised of a lab-on-chip module called Microarray which allows thousands of genetic samples to be scanned altogether. Microarray consists of a grid-structure of several spots, some or all having testing substance called probes. Each probe has DNA sequence sample of size of some Pico moles (around 10^{-12} moles), which hybridize with target DNA sample and with the help of fluorescent-label, presence of nucleic acid can be detected [1]. Detection procedure includes scanning of microarray in presence of laser by CCD camera. Thus image analysis of microarray slide conveys key role in analysis and detection procedures. Such images, may undergo some defects like blurring, irregular morphology of spots, partial viewport etc., causing analysis inaccurate [2]. In order to provide clear image for processing, image stabilization is the key factor after capturing the images as those images might have noises introduced because of shocks or non-uniform imaging systems. To discuss about correction of defects, prior information on definition of defect is required. In terms of microarray image processing, analysis is done based on two values which are spot number and intensity. The spot number is the index of spot, representing particular sample which is to be tested for analysis. After image capture, spots, if active, are represented using specific color (Intensity) that can be measured using color-scales like RGB or HSV or grayscale. Thus in order to analyze and diagnose accurately, these two values should be detectable to computer software. Considering image processing, each spot is represented by group of pixels in image which are having some morphology and intensity mapped between 0 and 255 for 8-bit images. Defects occurs when spot is undetectable or misinterpreted because its morphology and positioning is poorly represented or its intensity cannot be calculated. Thus in all relation to microarray images, defects can be seen as failure of detection of spot location and intensity.

II. PREVIOUS WORKS

As there is a very less research discussing the defects that can occurred, not many frameworks are available for defect analysis. One attempt is given by Peter Bajcsy [3], which in his paper, discussed the parameters of ideal microarray image and its representation in the form of morphology or illumination. The defects that violate the ideal nature of microarray spots are discussed to conclude correction algorithms and evaluating performances in terms of graphical as well as statistical aspects. Major factor discussed is variations among spot intensities, morphology, grid geometry and background that causes images unfit to be subjected for further analysis. This review also provides the ideal microarray image processing algorithm robust for above variations. Omar[4] reviewed four microarray image scanning methods that provide sequential image processing. Algorithms consisted segmentation procedure, calculations of background and foreground intensities and variations, finding local intensity and mapping outputs. Omar briefly discussed procedures followed in each algorithm along with their merits and demerits. Further, in conclusion, it suggested Edward's method comprising median blur removal filter and intensity subtractions. Jean [5], in his paper provided automatic image processing algorithm to provide segmentation of spots and restoration by analysing the defects. Segmentation is carried using watershed transform and then spots were detected using morphological transform. De-blurring procedure involved median blur. Performance is estimated by parameter of normalized mean intensity and compared with three standard algorithms provided by ScanAlize and GenePix.

III. IDEAL MICROARRAY IMAGE

In an ideal microarray image, every spot should be showing its intensity accurately and uniformly. Furthermore, few points should be considered while describing ideally analysable spot.

- Spot should retain original morphology in original grid structure in image.
- Image should not have any dye spills.
- All spots should remain uniform in size.
- All pixels in same spot should have equal intensity.
- All spots are clearly visible without any edge faults that can disturb the shape of spot.
- Image should be noise free, representing with clear background and foreground.

There are some other conditions to be satisfied for ideal image such as

- Image should take minimum space.
- It shall be analysable again without corrupting any original data.

But, for the sake of confined scope, these two conditions are not evaluated for ideal conditions.

IV. DEFECT ANALYSIS

Defect analysis is important step where possible defects are taken in consideration and characterized in order to provide suitable strategy to correct them. Mainly classification of defects in considered desired approach as if defects are classified into common category, it's easier to provide common solution for each group, rather than correcting them individually.

Defects are broadly classified into 3 categories:

1. Classification by nature of defect
2. Classification by cause of defect
3. Classification by impact factor

A. Classification by nature

These defects are classified by the characteristics they display. Characteristics are defining standards of classification by nature, based on physical appearance of the defect present in image. These are again further classified into 2 groups:

1. Intensity related defects

These defects display the wrong intensity values, clearly disturbing the uniform intensity spread across whole spot. Since the spot is not having uniform intensity, the values which are obtained need intensity extraction processing like median value algorithm etc.

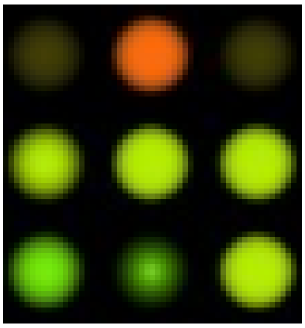

Fig.1 Intensity related defects

Notice how first spot is showing irregular intensity where majority of pixels are showing not uniform colour. Intensity is concentrated near centre of spot where as at edges, it decreasing gradually. This defect can cause inaccurate interpretation of intensity, leading to inaccurate analysis. Strategy to tackle these defects is by applying median intensity algorithm, which can detects the intensity by calculating median value present on spot.

2. Morphology related defects

Morphology deals with shape and structure of objects. Thus, morphological defects are defects which causes disturbance in spot morphology. One example of these defects can be shape transformation.

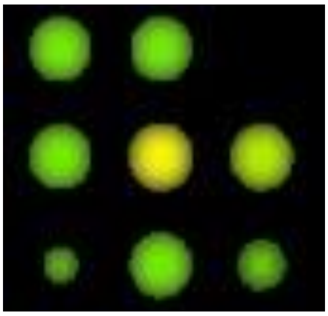

Fig.2 Morphology related defects

In above image, spots in bottom corners are represented by circles with different radii than actual spot. Such defects can lead to non-labelling spots as if spot size is less than minimum area for spot detection, it goes undetected.

These defects are corrected using adaptive circle detection where algorithms checks for multiple radii at once. After detecting the spot, it is reconstructed using spot reconstruction algorithms which draw circle of specific size over that spot.

B. Classification by cause

This classification is based upon the reason by which the defect is occurred. That root-cause analysis of defects turns out very effective for stabilization of microarray image processing algorithm. This classification, in reference of images, deals with environment through which defect has arrived. Hence these are further classified into environmental factors such as

1. Physical defects

Physical defects are defects which existed in faulty microarray pre-processing conditions. Best example of these type of defects is dye-spills, which gets perceived through camera and leads to inaccurate identification of spot morphology as well as spot intensity.

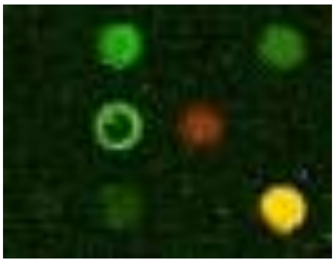

Fig.3 Physical defect

Above image displays the physical defects where spot is getting displayed as hollow circle instead of the filled one. This causes because of the physical issue or error, causing the material in the middle of spot getting washed out and hence its similarly reflected in image where the spot doesn't get identified using methods which detects spot as per minimum threshold area. These type of defects can be corrected in image construction step where only important details are collected and spot is drawn, overcoming the presence of defect.

2. Software defects

These defects are called noises. Defects occurring at software level reflects on the inefficiency for image processing software to provide "clean and noise-free" image. One example of this can be known as salt & pepper noise, which spreads randomly over image, disturbing the cleanliness of image. The above image displays random salt and pepper noise with blur. These type of noises can be corrected using bi-convolution method as Manjunath(2014) discussed. Convolution occurs after application of multiple filters like median filter or bilateral filter etc. After that, thresholding is applied to differentiate between background and foreground respectively.

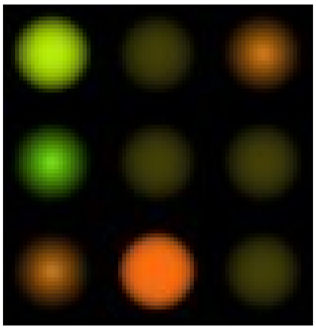

Fig.4 Blur

C. Classification by impact

This classification deals with defects in terms on impact they bare creating. This method of classification allows the priority preferences to carry major role in terms of correction. When a conflicting method of data identification is required, each defects is tested against its impact on both:

- Spot estimation, where spots cannot be accurately identified.
- Intensity extraction where intensity of the spot cannot accurately identified.

This classification provides three categories as follows:

1. Minimum impact
2. Medium impact
3. High impact

1. Minimum Impact

Minimum impact defects are defects whose presence have minimum impact on detection of spots and analysis. Thus, even when a defect is present in the image, if its impact factor is lower, images can be analyzed. Salt and pepper noises don't cause structural changes in image because anyhow spots can remain identifiable. Correction and stabilization of image don't correlate much with these defects in terms of time complexity.

2. Medium Impact

These kind of defects are having impact on analytical procedures, but these can be analyzed with certain degree manual adjustments like adjusting minimum area or threshold. Thus, once detected, these are defects which can be corrected using proper set of correction algorithms at an individual level considering spots. One prominent example of these kind of defects is irregular size of circles, among clusters of spots. This defect can be corrected by adjusting minimum area. Intensity related blurring is also among such defects where intensities can be corrected using thresholding.

3. High Impact

High impact defects are, as name suggests, defects with a scale that these disturb the analysis of the image as one or both parameters required are undetectable. These defects are imposed on either intensity factor or may disturbing the morphology of the spot or can have both impact making whole spot undetectable. Such defects are disastrous, impacting most of the cognition of the spot even human eye can't have. So such defects cannot even be corrected by human intervention, rendering whole scanning useless.

One example of these defects is displaced bio-wash material. This defect can cause inaccurate detection of spot and its intensity. Since it is a displacement, many spot detection algorithms can wrongly identify the spot location. If this defect occurs at initial spot locations, then whole grid can wrongly be created mis-fitting many spots. In addition to that, intensity is difficult to fetch because wrong estimation of spot location. This, is riskier way of dealing using algorithms (like automatic segmentation) using neural networks, with defective morphology as well as intensity extraction for the spots that define grid parameters. Dye spill is another defect that may inflict heavy impact on the spot and intensity extraction which ultimately results into either wrong analysis or unstable image for analysis.

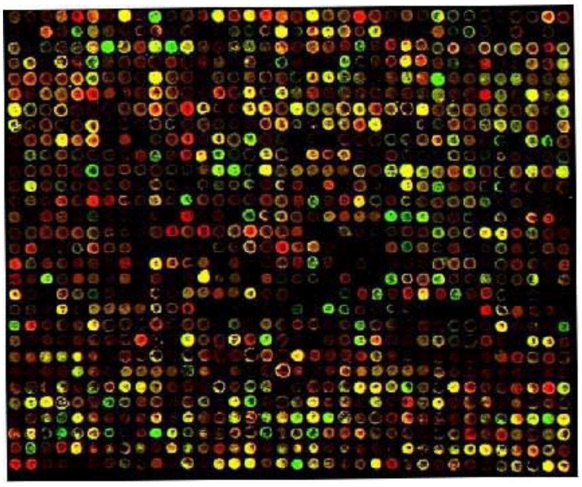

Fig.5 Image having medium impact defects

Considering the above image, one can notice the defects are variables of spot morphology and intensity thresholds. Thus correcting such image using image processing is easier by modifying preset processing parameters. Thus, this image stands as analyzable image after certain degree of image processing. Gridding approach also works well in case of segmenting image like Fig. 5. Yet, to some level of uncertainty, because of errors in detecting grid identifier spots, those algorithms failed in some cases. Thus, image with these impact factors are suitable for problem statements rather than heavy ones. Another example of such medium impact images is this image.

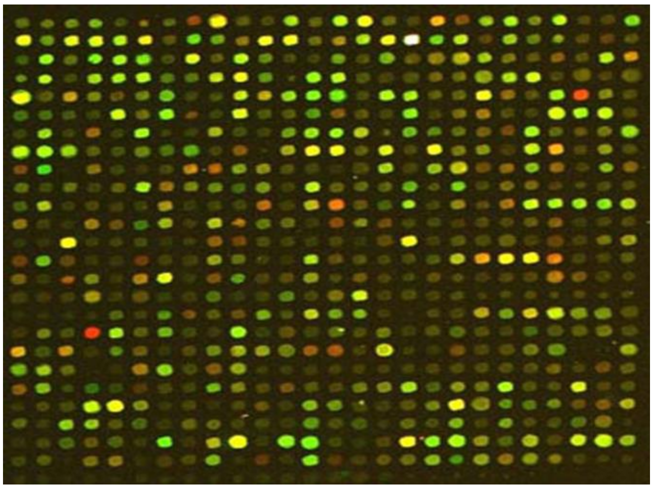

Fig.6 Image having medium impact defects

Notice how above image has irregular morphologies among spots and some spots have uneven spread of multiple intensities on same spot. Many clusters of white and dark pixels are visible over spots, which are noises. This image also proposes challenge for fully automatic algorithms with no pre-provisioned parameters. HCT algorithm also failed dramatically to locate spots which are not circular. Connected components algorithm fails in this image because white noises are creating virtual bridge among components which are nothing but spots. These components cannot be differentiated from one another because of such bridge. So it's always a challenging task to correct the defects which are having high impact factor.

V. Conclusion

This research proposes microarray image processing as a way of correcting the defects and preventing them to cause any disturbance in analysis. Hence, it provides the new way of understanding the defects occurring in microarray images by defining the ideal image and comparing the characteristics of both. Defect analysis classifies defects according to their nature, causes and impact. These three classifications allow clustering similar defects and application of suitable method to be applied for their correction. This defect analysis can be used as framework for designing algorithms to be used in microarray image processing.

References

[1] Taub, Floyd, Laboratory methods: Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. DNA. **2** (4): 309–327, 1983.

[2] Peter Bajcsy., DNA microarray image processing, 2005

[3] Peter Bajcsy, An Overview of DNA Microarray Image Requirements for Automated Processing in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) – Workshops, 2005.

[4] Omar Salem Baans, Analysis of Background Correction Methods for DNA Microarray Imaging, JERE, Volume 10, No.1 (27- 36), Special Issue on Bioprocess & Biosystem Engineering and their Downstream Processing, 2018.

[5] Jean Serra, Automatic analysis of DNA microarray images using mathematical morphology, in BIOINFORMATICS Vol. 19 no. 5 2003, pages 553–562, 2003.

[6] Zhu, Ningbo and Wang, Gang and Yang, Gaobo and Dai, Weiming, A fast 2d otsu thresholding algorithm based on improved histogram, in Pattern Recognition, 2009. CCPR 2009. Chinese Conference on: 1-5, 2009.

[7] Hamid Hassanpour, Najmech Samdiani, S. M. Mahdi Salehi, Using morphological transforms to enhance the contrast of medical images, by Egyptian journal of radiology and nuclear medicine, vol. 46, issue 2, 2015.

[8] W. Huber, M. Banejad, Anja Von Heydebreck, HolgerSultmann,A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression, by Bioinformatics, 2002;18 Suppl 1:S96-104.

[9] Ren-You Huang, Lan-Rong Dung, Cheng-Fa Chu, Yin-Yi Wu, Noise Removal and Contrast Enhancement for X-Ray Images, in Journal of Biomedical Engineering and Medical imaging, Vol. 3, No. 1, Feb 2016.