

Two Dimensional Gel Electrophoresis Image Inpainting for Improving Protein Spot Detection

Shaheera Rashwan

Informatics Research Institute, City of Scientific
Research and Technological Applications,
Alexandria, Egypt

Ashraf Khalil

Genetic Engineering and Biotechnology Research
Institute, City of Scientific Research and
Technological Applications, Alexandria, Egypt

ABSTRACT

Many specialized software programs are available for processing two dimensional gel electrophoresis (2-DGE) images. However, the anomalies existing in these images make the achievement of a reliable system for 2-DGE image analysis still difficult to reach. In this paper, we propose a new preprocessing technique applied to 2-DGE images. The new technique is based on image inpainting using Mumford Shah Euler Lagrange to discard anomalies such as vertical and horizontal streaking from these images. We also present a comparison of the analysis of 2-DGE images inpainted by the proposed technique and non-inpainted images using the known commercial software Delta2D. We compute the F-measure in both cases for three different 2-DGE images. The degree of improvement in F-measure reaches 18.5% in first image and 5.9% in second image and 3.8% in third image. Our new 2D Gel image preprocessing method based on Mumford Shah inpainting shows a significant improvement when comparing analysis of inpainted images with non-inpainted images.

Keywords

Two Dimensional Gel Electrophoresis (2-DGE), Image Inpainting, 2-DGE image preprocessing, Mumford Shah Inpainting Model.

1. INTRODUCTION

Two dimensional polyacrylamide gel electrophoresis (2-DE) technique is considered as one of the leading powers in the expansion of proteomics and protein studies. It provides the first step for further analysis of the differentially regulated protein spots using mass spectrometry [1]. The 2-DE has been used in many cases to disclose both physiological mechanisms and proteins with clinical pathologies that can help in the discovery of biomarkers.

2-DE is considered a powerful tool used for separation and fractionation of complex protein mixtures from tissues and cells since it enables separation of hundreds to thousands of proteins in one gel.

2-DE consists mainly of two dimensions of separation. In the first dimension, protein molecules are resolved depending on their isoelectric point (pI). In the second dimension, protein separation is performed based on molecular weight.

Therefore, the presence of different types and levels of noise frequently disturb 2-DGE image analysis. For example, the impulsive noise can spectrally happen in all frequencies when the sample is digitalized [14-15].

The background of 2-DGE images can also vary among samples, depending on the technical specifications of the imaging system used to capture the images. These variations

and anomalies often cause the complication of the analysis of 2-DGE images and affect the reproducibility of the results obtained. A preprocessing phase could be effective in reducing or eliminating these anomalies, thus reducing errors in the spot detection.

Cancelling this preprocessing stage may affect the results of the subsequent analysis since noise falsely identified as a protein, whereas real proteins of interest were missed.

There are three common objectives for the 2-DGE image preprocessing, i.e., pixel intensity regularization (image normalization), background correction, and noise reduction (filtration) [2].

In this paper, we present a new preprocessing method of 2D gel images based on image inpainting to discard vertical and horizontal streaking. The new method is based on Mumford-Shah inpainting algorithm and shows efficiency. The rest of the paper is organized as follows: Section 2 presents some of the related work. Section 3 presents the methodology. Section 4 presents the experimental results and the discussion. Section 5 concludes the work of the paper and we present some of the suggested future work.

2. RELATED WORK

Inpainting addresses the issues in filling the missing image parts, and it is designed owing to certain factors like, from eliminating scrapes in photos, reinstating sketches, and filling in the lost pixels of images caused by means of noisy channels. In the imaging field, the name “inpainting” refers to the repairing of images with loosed or missing data. It remains as a significant work in a variety of image reinstatement issues together with impulse noise removal, zooming, and scratch removal, and so on.

Liangtian He et al. [3] presented a novel wavelet frame dependent weighted reduction design for inpainting image.

Dai-Qiang Chen et al. [4] has proposed a novel dependent on Proximity Projection algorithm (PPA) based on the iteration of Newton.

Fei Wen et al. [5] presented a paper in 2017 on the basis of Alternative Direction Method of Multipliers (ADMM) technique. This paper represents the de-mixing and recuperation issues causing in signals existing in common vocabulary.

Vahid et al. [6] represented an efficient algorithm for non-texture inpainting of images based on the dominant orientation of local gradients. It introduces the concept called orientation matrix for faster and better inpainting.

Haixia Wang et al. [7] introduced a novel algorithm depending on space varying updating strategy for a corresponding

confidence parameter to enhance the substantial priority evaluation.

Zhihua Chen A et al. [8], suggested a novel dependent on Patch priority algorithm that offers local configuration multiplier to enclose adequate information according to the structure in the function of priority that makes sure about the configuration stability.

B. Patil and P. Patil [9] developed a novel inpainting model on the basis of Mumford Shah (MS) modeling after a pre-processing of the image done by Discrete Wavelet Transform (DWT). Then, the filter coefficient in DWT algorithm is optimized by Crow Search Algorithm (CSA). Smoothing image model is done using Reproducing Kernel Hilbert Smoothing (RKHS). The results in [9] show the efficiency of using Mumford Shah model in image inpainting and that it outperforms the other methods also the suggested new MS model by Patil and Patil causes a very small improvement over the conventional MS inpainting model. This justifies our choice to use the MS model in the 2D gel image inpainting method presented in this paper.

3. MATERIALS AND METHOD

The Mumford Shah method for image inpainting has been described in several papers [9-11] before. In this section, we present how we can use it as a preprocessing step to decrease the effect of horizontal artifacts in the 2D gel images.

The suggested method is as follows:

Step 1: We take a sample of the horizontal streaking as being the damaged part D of the input image u_{start}

Step 2: For a given grey value image u_{start} with inpainting domain (damaged part) D we want to reconstruct an image u by solving a minimisation problem iteratively via alternating solutions of the Euler-Lagrange equations

Step 3: The image u is the output image where most of horizontal streaking were discarded.

Step 4: Apply any commercial software of 2D gel image analysis (e.g. DELTA2D) to image u .

Figure 1 shows the steps of the new pre-processing method

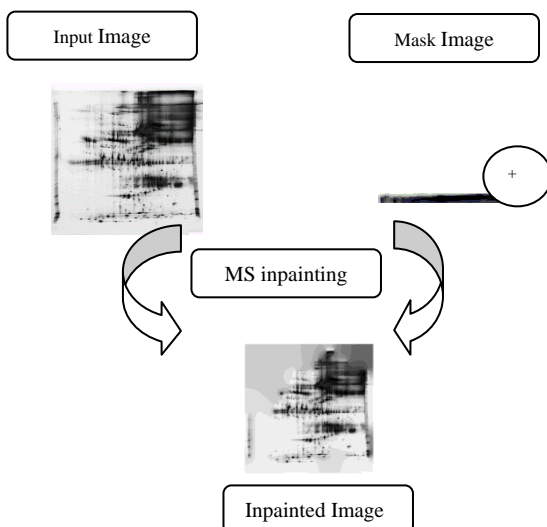


Figure 1. The Steps of our 2D gel Inpainting Model

4. RESULTS AND DISCUSSION

4.1. Data Used

The LECB 2-D PAGE [12] gel images database is available for public use. It contains data set of four types of experiments with over 300 gif images. It could be used for samples of several types of biological materials and for test data for 2D gel analysis software development and comparison with other similar samples. The LECB was the U.S. National Cancer Institute's Laboratory of Experimental and Computational Biology.

The four data sets are: Human leukemias, HL-60 cell lines, MOLT-4 cells and Fetal alcohol syndrome (FAS) - serum.

In our experiments, three images were chosen of FAS since they contain the horizontal streaking on which we focus in our study. Figure 2 shows the three datasets of FAS

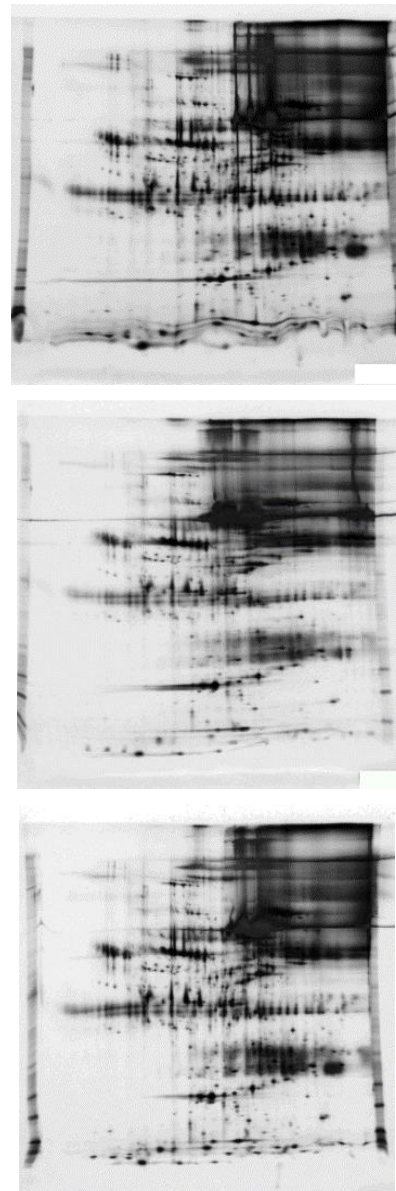


Figure 2. 2D gel images of the three data samples

4.2. Commercial Software

In this work, the commercial software Delta 2D were used [13] for 2D gel electrophoresis image analysis with and without our new pre-processing method based on image inpainting.

4.3. Quality Metrics for Evaluation

We evaluate the spot detection algorithm using the number of true positives (TP), false positives (FP) and false negatives (FN). TP is the number of spots detected correctly, FP is the number of regions mistakenly marked as spots, and FN is the number of spots not detected. There are also three quality metrics based on TP, FP and FN which are the sensitivity S and the precision P and the F-measure F.

$$S = \frac{TP}{TP+FN} \quad (1)$$

$$P = \frac{TP}{TP+FP} \quad (2)$$

$$F - measure = 2 \times \frac{S \times P}{S+P} \quad (3)$$

4.4. Experimental Results

From Tables 1 and 2, we can see that there is a significant improvement in the Precision (P) values in all three cases caused by a decrease in (FP) values i.e. the number of spots mistakenly marked as spots had decreased in the inpainted images.

We can also notice that the sensitivity (S) was not affected negatively by the inpainting process. The difference in the sensitivity values in the three cases is very low.

Table 1. Values of TP, FP, FN, sensitivity and precision of 2D images using Delta2D software without inpainting

Sample No.	TP	FP	FN	Sensitivity (S)	Precision (P)	
1	154	192	32	0.8279	0.4451	
2	105	211	24	0.8139	0.3323	
3	141	223	31	0.8198	0.3874	
Average						
					0.8206	0.3882
Standard Deviation						
					0.0070	0.0564

Table 2. Values of TP, FP, FN, sensitivity and precision of 2D images using Delta2D software with inpainting

Sample No.	TP	FP	FN	Sensitivity (S)	Precision (P)	
1	140	93	35	0.8	0.6009	
2	107	191	23	0.8231	0.3591	
3	139	201	30	0.8225	0.4088	
Average						
					0.8152	0.4562
Standard Deviation						
					0.0132	0.1277

From Table 3 and Figure 3, we can see that the F-measure had reached a high improvement of 18.53% in the first case and moderate improvements in the other two cases. This underlines the good effect of applying the inpainting pre-processing step.

Table 3. Values of F-measure using Delta2D software with and without inpainting

Sample no.	F-measure (Without Inpainting)	F-measure (With Inpainting)	% degree of improvement
1	0.5789	0.6863	18.5383
2	0.4719	0.5	5.9524
3	0.5261	0.5462	3.8108
Average	0.5257	0.5775	-
Standard Deviation	0.0535	0.0970	-

F-MEASURE

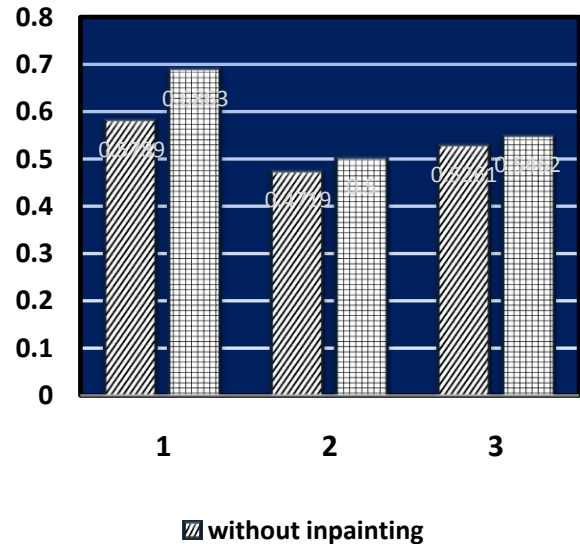


Figure 3. F-measure of spot detection in 2D gel images for the three samples with and without inpainting

In Figure 4, we notice that most of the horizontal streaking disappear in the inpainted images making the detection of incorrect spots rare to happen.

In sum, the new inpainting pre-processing model improves the detection of protein spots in 2D gel images in terms of the quality metrics Sensitivity (S), Precision (P), and F-measure. The added step to the 2D gel image analysis procedure decreases the negative effect of the contamination of the gel.

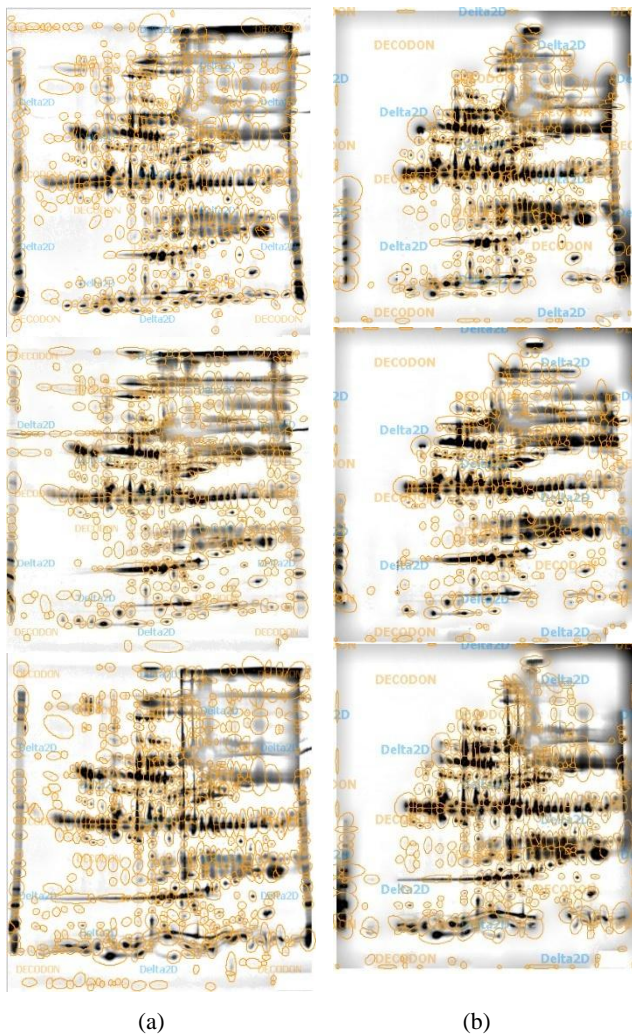


Figure 4. 2D gel images of the data samples analysed by Delta 2D software (a) without inpainting, and (b) with inpainting

5. CONCLUSIONS AND FUTURE WORK

In this paper, a new pre-processing method of 2D gel images using Mumford Shah inpainting model was presented. The new method had successfully discarded the horizontal streaking in the images which causes a more efficient analysis by the commercial software. There was a significant improvement in the F-measure calculated with and without inpainting. Our future work will focus on image inpainting for other artifacts in the 2D gel images.

6. REFERENCES

[1] Magdeldin, Sameh, Shymaa Enany, Yutaka Yoshida, Bo Xu, Ying Zhang, Zam Zureena, Ilambarthi Lokamani, Eishin Yaoita, and Tadashi Yamamoto. "Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis." *Clinical proteomics* 11, no. 1 (2014): 16.

[2] Goetz, Manuel Mauricio, Maria Constanza Torres-Madroñero, Sarah Röthlisberger, and Edilson Delgado-Trejos. "Preprocessing of 2-dimensional gel electrophoresis images applied to proteomic analysis: a review." *Genomics, proteomics & bioinformatics* 16, no. 1 (2018): 63-72.

[3] He L, Wang Y. Iterative support detection-based split bregman method for wavelet frame-based image inpainting. *IEEE Transactions on Image Processing*. 2014 Dec;23(12):5470-85.

[4] Chen DQ, Zhou Y. Inexact alternating direction method based on proximity projection operator for image inpainting in wavelet domain. *Neurocomputing*. 2016 May 12;189:145-59.

[5] Wen F, Adhikari L, Pei L, Marcia RF, Liu P, Qiu RC. Nonconvex regularization-based sparse recovery and demixing with application to color image inpainting. *IEEE Access*. 2017;5:11513-27.

[6] Alilou VK, Yaghmaee F. Non-texture image inpainting using histogram of oriented gradients. *Journal of Visual Communication and Image Representation*. 2017 Oct 1;48:43-53.

[7] Wang H, Jiang L, Liang R, Li XX. Exemplar-based image inpainting using structure consistent patch matching. *Neurocomputing*. 2017 Dec 20;269:90-6.

[8] Chen Z, Dai C, Jiang L, Sheng B, Zhang J, Lin W, Yuan Y. Structure-aware image inpainting using patch scale optimization. *Journal of Visual Communication and Image Representation*. 2016 Oct 1;40:312-23.

[9] Patil BH, Patil PM. Crow search algorithm with discrete wavelet transform to aid Mumford Shah inpainting model. *Evolutionary Intelligence*. 2018 Oct 1;11(1-2):73-87.

[10] Shen J. Inpainting and the fundamental problem of image processing. *SIAM news*. 2003 Jun;36(5):1-4.

[11] Esedoglu S, Shen J. Digital inpainting based on the Mumford–Shah–Euler image model. *European Journal of Applied Mathematics*. 2002 Aug;13(4):353-70.

[12] 2D Gel databases <http://bioinformatics.org/lecb2dgeldb/> (last accessed 5-5-2019)

[13] DECODON Delta 2D <https://www.decodon.com/delta2d.html> (last accessed 5-5-2019)

[14] Faheem, Muhamed Talaat, Shaheera Rashwan, Amany Sarhan, and Bayumy A. Youssef. "De-noising 2D gel images using wavelet transform based on genetic algorithm." In *Proc 10th WSEAS Int Conf Appl Comput Appl Comput Sci*, pp. 187-94. 2011.

[15] Rashwan, Shaheera, Talaat Faheem, Amany Sarhan, and Bayumy AB Youssef. "A relational fuzzy c-means algorithm for detecting protein spots in two-dimensional gel images." In *Advances in Computational Biology*, pp. 215-227. Springer, New York, NY, 2010.