Fuzzy Watershed Algorithm: An enhanced algorithm for 2D gel electrophoresis image segmentation

Shaheera Rashwan[1], Amany Sarhan[2], Muhamed Talaat Faheem[3], Bayumy.A.Youssef[1]

[1]*Informatics Research Institute, City for Science and Technology, Borg ElArab, Alexandria, Egypt*
[2]*Computer science and Automatic Control Engineering Department, Faculty of Engineeing, University of Tanta, Tanta, Egypt*
[3]*EE dept-Faculty of Engineering-Taif University, KSA*

Abstract

An important issue in the analysis of two-dimensional electrophoresis images is the detection and quantification of protein spots. The main challenges in the segmentation of 2DGE images are to separate overlapping protein spots correctly and to find the abundance of weak protein spots. To enable comparison of protein patterns between different samples, it is necessary to match the patterns so that homologous spots are identified. In this paper, we describe a new robust technique to segment and model the different spots present in the gels. The Watershed segmentation algorithm is modified to handle the problem of over segmentation by initially partitioning the image to mosaic regions using the composition of fuzzy relations. The experimental results showed the effectiveness of the proposed algorithm to overcome the over segmentation problem associated with the available algorithm. We also use a wavelet denoising function to enhance the quality of the segmented image. The parameters of the wavelet function are obtained using the Genetic Algorithm search technique. The results of using the denoising function before the proposed Fuzzy Watershed segmentation algorithm is very promising as they are better than those without denoising.

Keywords: Protein Spot Detection, Watershed Segmentation, over-segmentation, Fuzzy Relations

1. Introduction

Two-dimensional gel electrophoresis (2-D Gel) enables separation of mixtures of proteins due to differences in their isoelectric points (pI), in the first dimension, and subsequently by their molecular weight (MWt) in the second dimension. Other techniques for protein separation exist, but currently 2-D Gel provides the highest resolution allowing thousands of proteins to be separated. The great advantage of this technique is that it enables, from very small amounts of material, the investigation of the protein expression for thousands of proteins simultaneously.

In this paper, the most important issues and challenges related to digital image analysis of the gel images will be addressed, namely the segmentation of the images. The watershed algorithm was used to segment the two-dimensional electrophoresis gel (2-D Gel) images. The watershed algorithm [1,2,10,11] is very well suited for the problem of segmenting the different spots in a 2-D gel images, because after applying a small mean-filter, these spots are characterized by a monotonic increasing and thereafter decreasing shape. In this way it is possible to detect the catchment basins belonging to the different gel spots.

This is a very robust approach: varying background intensity has no influence on the finding of the different spot regions. To exclude small regions corresponding to background noise, a

threshold was chosen for the minimal size of the basins. The remaining basins delineate the regions of most spots. However, some spots overlap in such a way that they give rise to only one catchment basin, and as a result they will be identified as one spot. To segment the spots from the background, the density peaks in the image have to be found. A big advantage of this algorithm is that it is robust in the sense that it is not influenced by a variable background (low-frequency variations). Unfortunately the watershed algorithm suffers from the over segmentation problem. In this paper, in order to overcome such problem, we propose the use of fuzzy notion to the original algorithm. The watershed algorithm will be preceded by partitioning step to the image to convert it to mosaic regions using the composition of fuzzy relations. Then the watershed algorithm is applied to the produced mosaic image. The experimental results run on a group of protein gel images showed the effectiveness of the proposed algorithm to overcome the over segmentation problem available in the original watershed algorithm.

This paper is organized as follows: section 1 presents the introduction. Section 2 summarizes the background. Section 3 summarizes the related work. Section 4 introduces the proposed watershed algorithm using the composition of fuzzy relation. Section 5 shows the software results of the proposed algorithm. Section 6 presents an application of wavelet denoising on images before segmentation. Section 7 concludes and discusses the software results of the proposed algorithm. Finally a list of references is given.

2. Backgrounds

2.1 The Watershed Algorithm

The watershed algorithm is a very robust for detecting spots, with the major advantage that there is no need for a background subtraction. Regarding this, the major disadvantage of the algorithm which is the over segmentation must be overcome.

The Watershed segmentation is a technique developed from morphological algorithms, which follows a geological analogy. The image to be segmented can be considered as a topographical surface, S, where the gray levels or image intensities, $I(x,y) = I(s)$ correspond to altitude values [14].

A minimum at an altitude value j, m^j, in this landscape, is a dip in the ground surrounded by strictly higher land. A catchment basin, $CB_i(m_i^j)$, is then the area around the minimum $m_i^{\ j}$ in S where water falling on it would flow down into the minimum.

At each pixel where two or more catchment basins meet, an imaginary 'dam' is built. At the end of a recursive process, each minimum is surrounded by dams, which delimit the associated catchment basins. These dams correspond to the watersheds of the topographical surface WT(S). This type of morphological transform can also be seen as an edge detector as it can naturally identify boundaries of objects within an image.

Image data may be interpreted as a topographic surface where the gradient image gray-levels represent altitudes. Region edges correspond to high watersheds and low-gradient region interiors correspond to catchment basins.

Catchment basins of the topographic surface are homogeneous in the sense that all pixels belonging to the same catchment basin are connected with the basin's region of minimum altitude (gray-level) by a

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

563

simple path of pixels that have monotonically decreasing altitude (gray-level) along the path. Such catchment basins then represent the regions of the segmented image.

Briefly explained, the algorithm can be divided into three phases: Firstly, all pixels in the gradient image G(I) are scanned looking for regional minima. Let us define N, the set of neighbors, (x', y'), for a pixel (x,y) in G(I). When 8-connectivity is used, x' = {x - 1, x, x + 1}; y' = {y- 1, y, y + 1}. If G(x,y) > G (x', y') x', y' ∈ N(x,y), then G(x,y) is labeled as non-regional-minima (NRM) and put into a first-input–first-output (FIFO) queue Q.

Subsequently, while Q is not empty, its first element is popped out. Let G (x', y') be the first output element of Q. If the label of G(x'', y'') is void, $x'', y'' ∈ N(x', y')$ and G(x,y) = G(x'', y''), then the label G(x'', y'') is set to NRM and G(x'', y'') is put in Q.

In a second step, the adjacent pixels of the minima found are put into an ordered queue (OQ). Starting from label i = 1, all pixels in G(I) are scanned again. If the label of G(x,y) is void, then G(x,y) ∈ CB_i and G(x,y) is put in a FIFO queue Q. Again, while Q is not empty, its first element is popped out.

Let G (x', y') be the first output element of Q. If the label of G(x'', y'') is void, $x'', y'' ∈ N (x', y')$, then G(x'', y'') ∈ CB_i and G(x'', y'') is put in Q; otherwise G(x'', y'') is labeled NRM and put in a gray value ordered queue OQ.

In the final stage, pixels in the ordered queue with the lowest gray value are popped out. Let G(x,y) be the

first output element of OQ. If label of G(x', y') is void, $x', y' ∈$ N(x,y), then G(x,y) ∈ CB_k if G$(x', y') ∈ CB_k$ for k = 1,. . . ,i.

2.2 The compositions of fuzzy relations

2.2.1 Fuzzy Relations

A Fuzzy relation generalizes classical relation into one that allows partial membership and describes a relationship that holds between two or more objects.

Example: a fuzzy relation "Friend" describes the degree of friendship between two persons (in contrast to either being friend or not being friend in classical relation!)

A fuzzy relation R is a mapping from the Cartesian space X x Y to the interval [0,1], where the strength of the mapping is expressed by the membership function of the relation $\mu_R(x, y)$

$$\mu_R : A \times B \to [0,1]$$

$$R = \{((x, y), \mu_R(x, y) | \mu_R(x, y) \geq 0, x \in A, y \in B\}$$
(1)

2.2.2 The max- min composition of Fuzzy Relations

Two fuzzy relations *R* and *S* are defined on sets *A*, *B* and *C*. That is, $R \subseteq A \times B$, $S \subseteq B \times C$. The composition $S \bullet R = SR$ of two relations *R* and *S* is expressed by the relation from *A* to *C*:

For $(x, y) \in A \times B$, $(y, z) \in B \times C$,

$$\mu_{S \bullet R}(x, z) = \max_y [\min(\mu_R(x, y), \mu_S(y, z))]$$
(2)

$$= \bigvee_y [(\mu_R(x, y) \wedge \mu_S(y, z))]$$
(3)

MS · R = MR · MS
(matrix notation) (max-min composition)

2.2.3 The max- Product composition of Fuzzy Relations

Two fuzzy relations R and S are defined on sets A, B and C. That is, $R \subseteq A \times B$, $S \subseteq B \times C$. The composition $S \bullet R = SR$ of two relations R and S is expressed by the relation from A to C:

For $(x, y) \in A \times B$, $(y, z) \in B \times C$,

$$\mu_{S \bullet R}(x, z) = \max_{y}[\mu_R(x, y) \cdot \mu_S(y, z)]$$

(4)

$$= \bigvee_{y}[(\mu_R(x, y) . \mu_S(y, z))]$$

(5)

$$MS \cdot R = MR \cdot MS$$

(matrix notation) (max-product composition)

3. Related Work

In [12], Hoang et al. presented a novel approach for protein spot detection, which is a marker-free Watershed that does not require specification of predefined markers for the process of finding watershed contour lines. This approach includes a selective nonlinear filter and pixel intensity distribution analysis for removing local minima which causes over-segmentation when applying watershed transform. It then superimposes those true minima over the reconstructed gradient image before applying Watershed transform for spot segmentation. The effectiveness of this marker-free approach was experimentally comparable with other methods.

In [13], Lin and kuo have developed an adaptive mechanism to adjust the level of detail and determine the threshold value of watershed. The over-segmentation drawback is overcome by applying directed graph version of watershed transform

algorithm and morphological opening operation. Labelling and region growing techniques were adapted to extracted individual spots features.

In [14], the watershed algorithm was used for spots segmentation in 2DGE images. But the paper is more focused on using the diffusion principle in modelling the spots. In [15], marker-based watershed segmentation methods were used to improve the segmentation of the protein spots from the varying background. In our work, we will introduce the notion of fuzzy relations to handle the problem of over-segmentation often produced by the watershed algorithm.

4. The proposed Watershed algorithm using the composition of Fuzzy Relations

In the proposed algorithm, we intend to add a new phase before applying the Watershed algorithm. This phase is considered as a preparation phase that transforms the image into mosaic image using the notion of the composition of fuzzy relations. We call the new algorithm fuzzy watershed segmentation (FWS) algorithm.

In our work, we will use two relations: $R_1 \circ R_2$. The advantage of using only two relations is the ease of the algorithm and the elimination of redundancy. Moreover, the connectivity is an important parameter for the watershed algorithm and adjusting this parameter before applying the algorithm has the intention of improving the algorithm by simplifying the original image to a mosaic image and reducing the error. The two relations are defined as follows:

R_1: x_i has 3×3 neighborhood x_j,

R_2: x_j has gray value y such that y belongs to cluster Z.

Figure 1 shows the relation R_1 as the neighborhood of point x_i.

x_j	x_j	x_j
x_j	x_i	x_j
x_j	x_j	x_j

Figure 1: The 3×3 neighborhood pixels

R_1 is a fuzzy relation defined as follows:

$$\mu_{R_1}(x_i, x_j) = |gv(x_i) - gv(x_j)|$$
(6)

Where xj is a 3x3 neighborhood of xi Z is the set of clusters initially partitioned as follows: $Z_1, Z_2, ..., Z_n$. Each cluster contains 256/n points defined as Z^i from $i*256/n$ to $(i+1)*256/n$ where i takes the values 0,1,…,n and 256 is the number of gray level values in the 2D gel electrophoresis image. In the experiments of the following section (section 5), n takes the value 128 which means reducing the original image to "half" of the available gray value levels, which yields to more simplified mosaic images.

R_2 is a crisp relation defined as follows:

$$\mu_{R_2}(x_j, Z) = \begin{cases} 1 & if \ gv(x_i) \in Z \\ 0 & elsewhere \end{cases}$$
(7)

The whole procedure of Watershed simplification can be reduced to the application of the following composition rule:

R1 ∘ R2: x_i is connected to x_j and x_j belongs to cluster Z.

In our approach, there is no need to apply the Fuzzy C-means algorithm as in Patino [3] since it is overhead to do clustering before the Watershed. The proposed partitioning method is much easier and reduces the complexity of the algorithm and the labelling is taken by maximizing the degree of membership values over all clusters, i.e,

$$x_{new} = \max_{Z_i = Z_1}^{Z_n} \mu_{R1 \circ R2}(x_{old}, z_i)$$
(8)

Since the second relation R_2 is a crisp relation, then the max-min composition is equivalent to the max-product. After partitioning the original image to mosaic regions, the watershed algorithm can be applied to the simplified image and hence reduces the over-segmentation problem.

We can present the proposed algorithm as in the following steps:

Fuzzy Watershed Based Algorithm (original image, noclstrs)

Output: segmented image

1: Initialize clusters Z as equal size of partitions over 256 gray level value in the image- where number of partitions is defined by noclstrs

2: For each pixel, find

$$\min_{x_j \, neighbor\, x_i} \; (\mu_{R_1}(x_i, x_j), \mu_{R_2}(x_j, Z))$$

where $\mu_{R_1}(x_i, x_j) = \left| gv(x_i) - gv(x_j) \right|$ **and**

$$\mu_{R_2}(x_j, Z) = \begin{cases} 1 & if \;\; gv(x_i) \in Z \\ 0 & elsewhere \end{cases}$$

3: Label pixels by applying the composition of fuzzy relation:

$$x_{new} = \max_{Z_i = Z_1}^{Z_n} \mu_{R1 \circ R2}(x_{old}, z_i)$$

4: Apply the Watershed algorithm to the resulted mosaic image

The following figure (Figure 2) shows an example of the application of the FWS algorithm steps on a 2D gel image.

5. Experimental results

The LECB 2-D PAGE gel images database is available for public use [16,17]. It contains data sets from four types of experiments with over 300 gif images. We choose randomly seven data samples to study the effect of applying the new algorithm versus the original watershed algorithm. The first three data samples are human Leukemia data samples then the next two data samples are blood lymphocytes and the last two data samples are for fetal alcohol syndrome. An example of one of these samples is shown in figure 3.
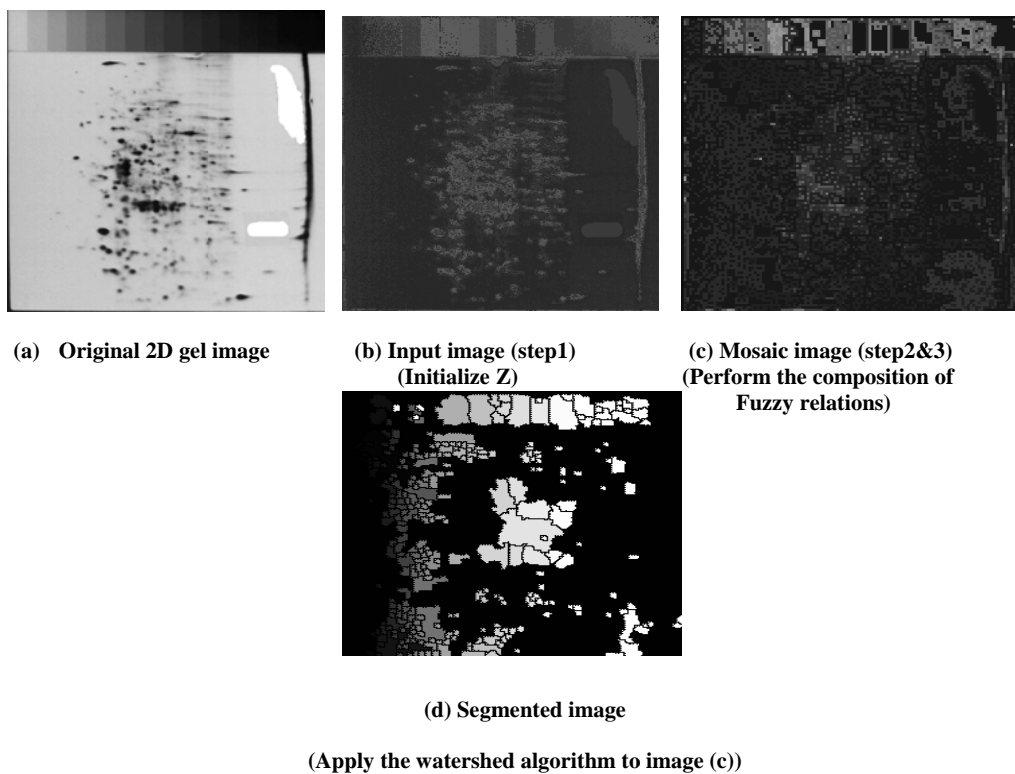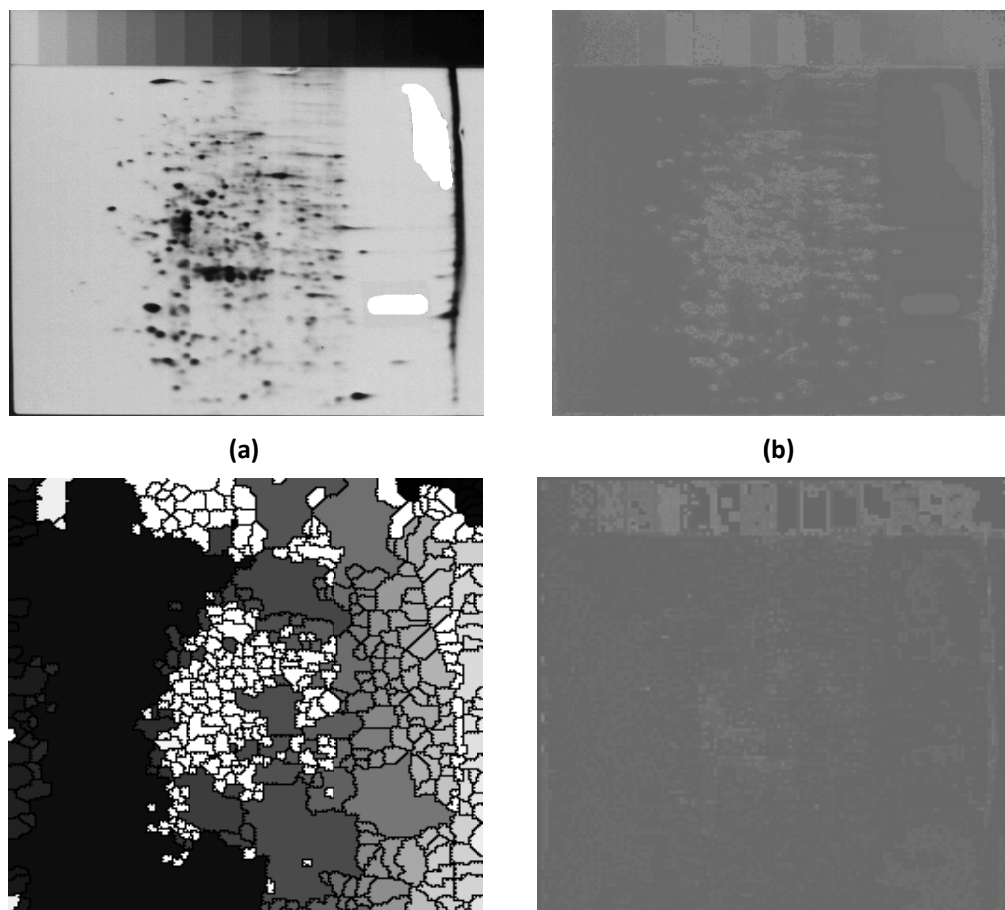
IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

567

(a) Original 2D gel image **(b) Input image (step1)** **(c) Mosaic image (step2&3)**
 (Initialize Z) **(Perform the composition of**
 Fuzzy relations)

(d) Segmented image

(Apply the watershed algorithm to image (c))

Figure 2: Applying the steps of the Fuzzy Watershed Segmentation algorithm on a 2DGE image

(a) **(b)**

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
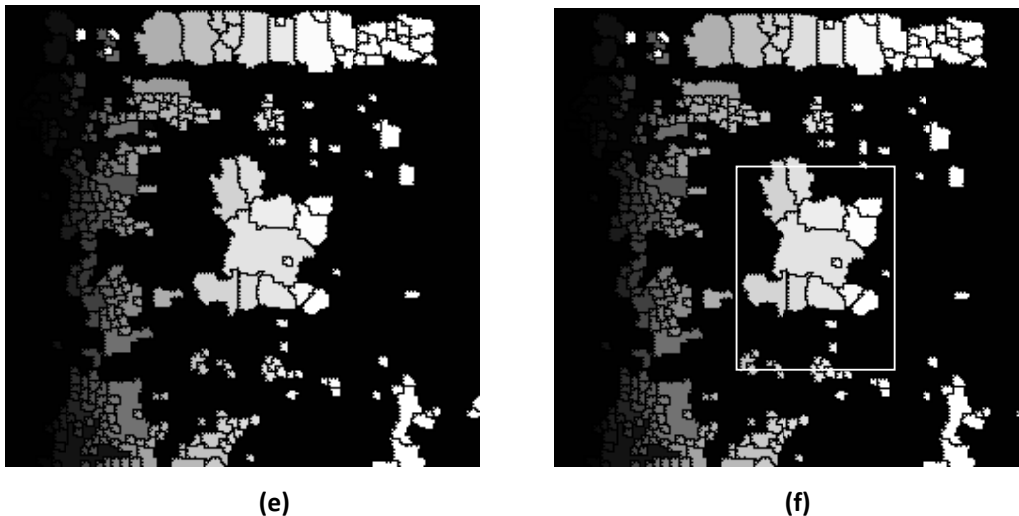ISSN (Online): 1694-0814
www.IJCSI.org

568

(e)

(f)

Figure 3: 2-D gel electrophoresis image of Patient- Human leukemias (a) original image, (b) gradient image, (c) gradient image after applying the Watershed algorithm, (d) mosaic simplified image of the gradient image after applying composition of fuzzy relations, (e) Gradient image after applying FWS, and (f) detected clusters after applying FWS

In our work, we used the evaluation method E_{CW} [18] for evaluating the new segmentation algorithm because it identifies the degree of under-segmentation –case of spots exist and undetected by the segmentation algorithm, and the degree of over-segmentation- Case of spots don't exist and falsely detected by the segmentation algorithm.

Evaluation method E_{CW} computes the intra-region color error, $E_{in.tra}$, as the proportion of misclassified pixels in an image. A misclassified pixel is defined as a pixel whose color error (in L*a*b space) between its original color and the average color of its region is higher than a pre-defined threshold.

1) Eintra of E_{CW}

$$= \frac{\sum_{p \in I} \mu(\left\|C_j \quad \text{Eintra} \quad C_x^s(p)\right\|_{L*a*b} - TH)}{S_I}$$

(9)

where $C_x^o(p)$ and $C_x^s(p)$ are pixel feature value (color components in CIE L*a*b space) for pixel p on original and segmented image, respectively, TH is the threshold to judge significant difference, and μ(t) = 1 when t > 0, otherwise μ(t) = 0.

2) Einter of E_{CW}

$$= \sum_{i=1}^{N} \sum_{j=1, j \neq}^{N} \text{Einter} \quad - \left\|C_x^o(p) - C_x^s(p)\right\|_{L*a*b}$$

(10)

where w_{ij} denotes the jointed length between R_i and R_j, TH is the threshold to judge significant difference, and Z is a normalization factor. S_I is the number of data samples in the image

In this section, we will use the E_{CW} evaluation error to evaluate the performance of the new algorithm the Fuzzy Watershed Segmentation algorithm (FWS) versus the original watershed algorithm (WS). As we are

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

569

interested in improving the over-segmentation without a serious declination in the under segmentation, we begin our comparison in this section by **Einter** which measures the over-segmentation error. Then, we will also measure the under-segmentation error (**Eintra**) to ensure that its values are not hardly affected by the improvement in the **Einter**. For estimating the values of TH and Z, we used ground truth images to know appropriate error of the segmentation as in [18]. We set the threshold TH = 10 and the normalization factor Z = 1000000.
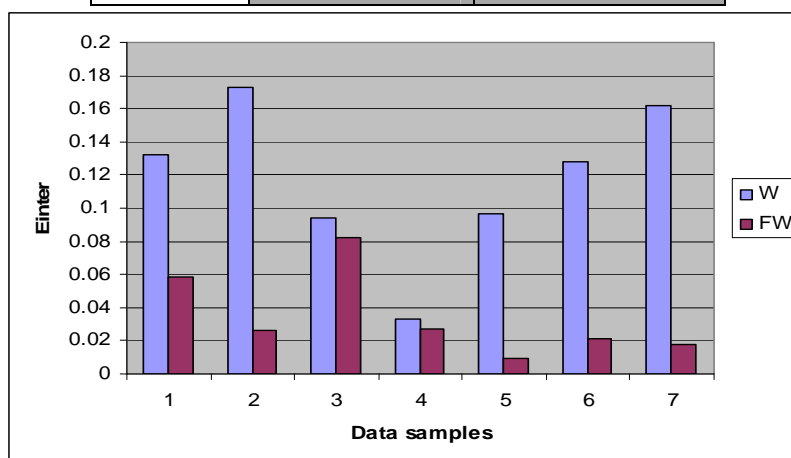
Table 1: The inter region error (Einter) of the watershed algorithm and the Fuzzy watershed algorithm on seven data samples

Data Sample no	Einter of watershed algorithm	Einter of Fuzzy watershed algorithm
1	0.1323	0.05867
2	0.17279	0.02635
3	0.09446	0.08236
4	0.03317	0.02723
5	0.09691	0.00921
6	0.1279	0.02121
7	0.16179	0.01788

Figure 4: The inter region error (Einter) of the watershed algorithm and the Fuzzy watershed algorithm on seven data samples

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
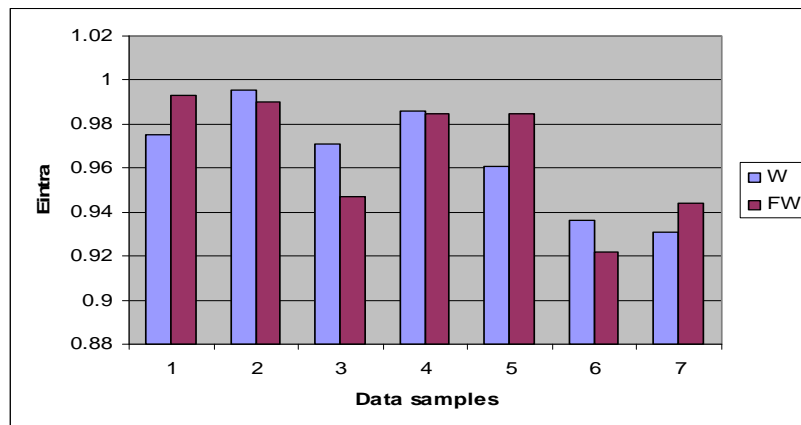ISSN (Online): 1694-0814
www.IJCSI.org

570

Table 2: The intra region error (Eintra) of the watershed algorithm and the Fuzzy watershed algorithm on seven data samples

Data Sample no.	Eintra of watershed algorithm	Eintra of Fuzzy watershed algorithm
1	0.97531	0.9929
2	0.99524	0.98994
3	0.97095	0.94678
4	0.9857	0.98464
5	0.96063	0.98448
6	0.93622	0.92159
7	0.93098	0.94414

Figure 5: The intra region error (Eintra) of the watershed algorithm and the Fuzzy watershed algorithm on seven data samples

The shaded cells in the two tables 1 and 2 represent the improvement caused by the FWS algorithm versus the original watershed algorithm. In other words, the shaded cells are when **Einter** and the **Eintra** decrease when applying the Fuzzy watershed based algorithm versus applying the original watershed algorithm. Notice that according to the **Einter** evaluation metric, the FWS algorithm had reduced the over-segmentation in all seven cases (100% success) and the improvement reached **14%** in data samples 2 and 7 .

Moreover, the **Eintra** evaluation metric, the FWS algorithm enhanced the results on 4 data samples from the 7 data samples (57% of the samples) with an average of 3% improvement in minimizing error for the 4 samples. In case of non-improvement in the other 3 samples, the difference was 2% in the worst case (data sample 5) which means that the proposed algorithm was

able to detect protein spots more precisely than the watershed algorithm.

We can also observe that in the first five data samples, which are data for the human leukemia and human blood lymphocytes, where exist the problems of ghost (weak) spots and noisy background, the proposed algorithm, compared to the original Watershed algorithm, succeeded in reducing the problem of over-segmentation but fails in identifying weak spots as in samples 1 and 5.

For the last two data samples, which are the Fetal Alcohol Syndrome, where exist the problems of contamination of gels and overlapped spots, the proposed algorithm, compared to the original Watershed algorithm, succeeded in reducing the problem of over-segmentation but fails in identifying overlapped spots as in sample 7.

6. Application of wavelet denoising based on genetic algorithm

Removal of noise is important step in order to obtain more accurate data, to allow for automatic analysis in high throughput proteomics and to understand software limitations.

The denoising methods commonly used so far, have the tendency to deform the protein spots on the gel to the extent that they create extraneous spots i.e. artifacts. This is a serious problem since insufficient or improper denoising affects the whole image processing pipeline from its early stages. So, it impacts negatively all the subsequent processes, such as spot detection, spot quantification, as well as spot matching across gels.

In their paper, Soggiu et al [21] used the undecimated (redundant) discrete wavelet transform to de-noise the 2Dgel images. They justified their choice of this form of wavelet by their saying that dealing with more complex data settings might involve non-orthogonality and the need to shift to non-decimated (or stationary) wavelet transforms (*ndWt*). By experimenting with different quantile values, they were able to interactively explore the best threshold for the given application. They reported evidence of quantile thresholding with variable accuracy levels (0.85, 0.99) for health-disease sample comparisons, and between diseased samples.

We use the genetic algorithm (GA) for adjusting the parameters of the wavelet function used in the denoising of the images, i.e, finding the best string that maximizes PSNR. In our work, we use the peak signal to noise ratio (PSNR) as fitness function, and use GA operators, such as selection, crossover, mutation, etc., to optimize the parameters of wavelet transform to improve denoising performance.

The denoising function we are going to use is the matlab function (**wden**) which is a one-dimensional discrete orthogonal wavelet transform function that performs an automatic denoising process of a one-dimensional signal using wavelets. This orthogonal function has 4 parameters which are: TPTR, SORH, SCAL, and N with each one having some possibilities as follows:

1- TPTR: a string that contains the threshold selection rule: 'heursure' which is a heuristic variant of the Stein' Unbiased Risk Estimation (SURE)[20], and 'minimaxi' for minimax thresholding which uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics in order to design estimators. Since the denoised signal can be assimilated to the estimator of the unknown

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

572

regression function, the minimax estimator is the one that realizes the minimum of the maximum mean square error obtained for the worst function in a given set. We need one bit for this value where (0) represents 'heursure' and (1) represents 'minimaxi'.

2- SORH: ('s' or 'h') is for soft or hard thresholding where the hard thresholding operator is defined as:

$$D(U, \lambda) = U \text{ for all } |U| > \lambda \quad (3.15)$$

Hard threshold is a "keep or kill" procedure and is more intuitively appealing.

The soft thresholding operator is defined as

$$D(U, \lambda) = sgn(U)max(0, |U| - \lambda) \quad (3.16)$$

Soft thresholding shrinks coefficients above the threshold in absolute value.

We need one bit for this value where (0) represents soft thresholding and (1) represents hard thresholding.

3- SCAL: defines the multiplicative threshold rescaling: 'one' for no rescaling and 'sln' (01) for rescaling using a single estimation of level noise based on first-level coefficients and 'mln' (10 or 11) for rescaling using level-dependent estimation of level noise.

Wavelets can be realized by iteration of filters with rescaling. The DWT is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in Figure 6.
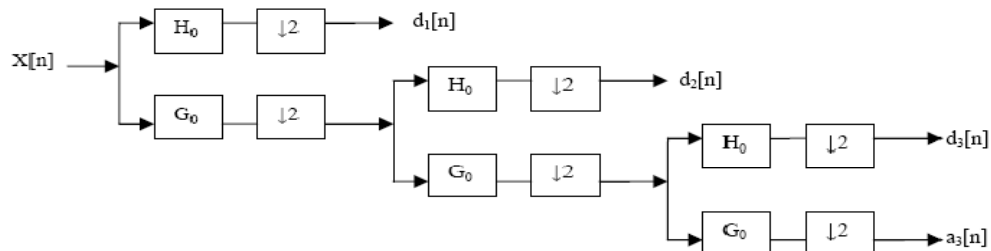

Figure 6: Three-level wavelet decomposition tree.

This is called the Mallat algorithm [19] or Mallat-tree decomposition. Its significance is in the manner it connects the continuous time mutiresolution to discrete-time filters. In the figure, the signal is denoted by the sequence x[n], where n is an integer. The low pass filter is denoted by G0 while the high pass filter is denoted by H0. At each level, the high pass filter produces detail information; d[n], while the low pass filter associated with scaling function produces coarse approximations, a[n].

We need two bits for this value where (00) represents no rescaling, (01) represents rescaling using a single estimation, (10) and (11) represent rescaling using level-dependent estimation of level noise.

4- N: Wavelet decomposition is performed at level N such that N=8. We need 3 bits (000 represents level 1 to 111 level 8).

So, the string of the GA chromosome will be of length 7 bits organized as shown in Figure 7. The

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

573

figure also shows the crossovering step in GA. If we want to do the octal encoding, we must perform the wavelet decomposition at level N such that N=16 for 4 bits string (0000 Lev 1 to 1111 Lev 16).
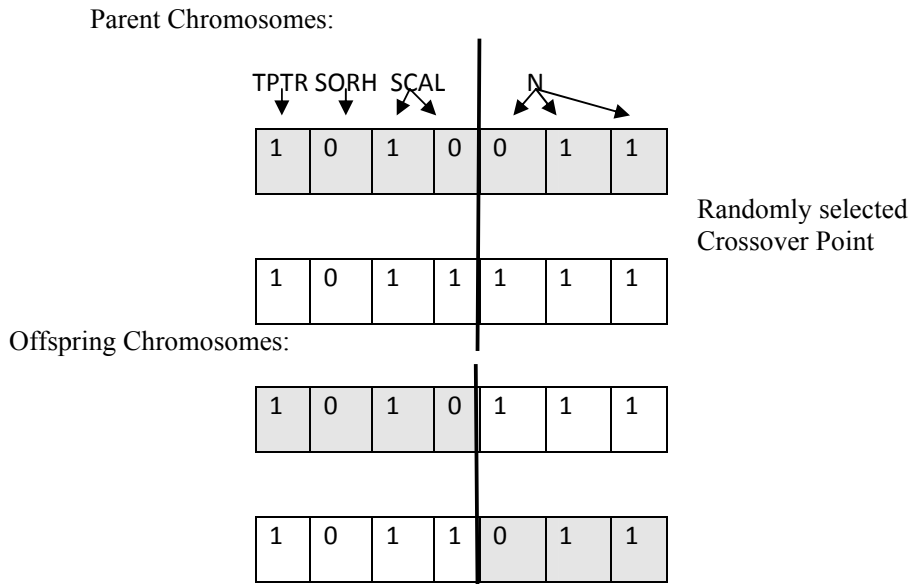
Parent Chromosomes:

Offspring Chromosomes:

Figure 7: The chromosomes of the genetic algorithm used in the denoising step

We applied the genetic algorithm to find the best string that maximizes the image PSNR (fitness function chosen). The initial population was generated randomly of size 50. The algorithm iterates 200 iterations. The best obtained string is (0110001) which means that: the value TPTR is heursure, the value SORH is hard, the value SCAL is mln and N (decomposition level) = 2.
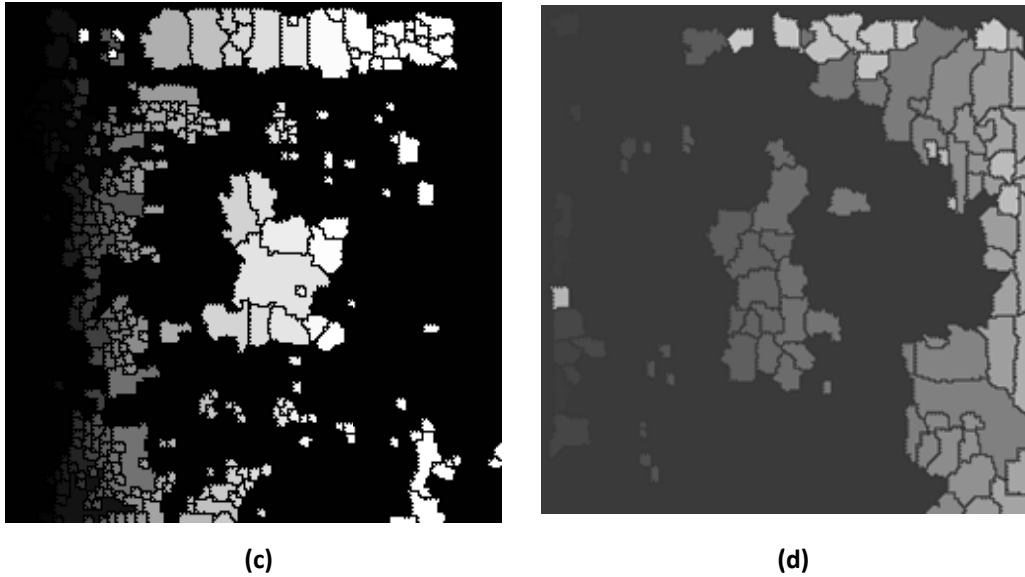
5.1 Numerical evaluation results

(a)   (b)

(c)   **(d)**
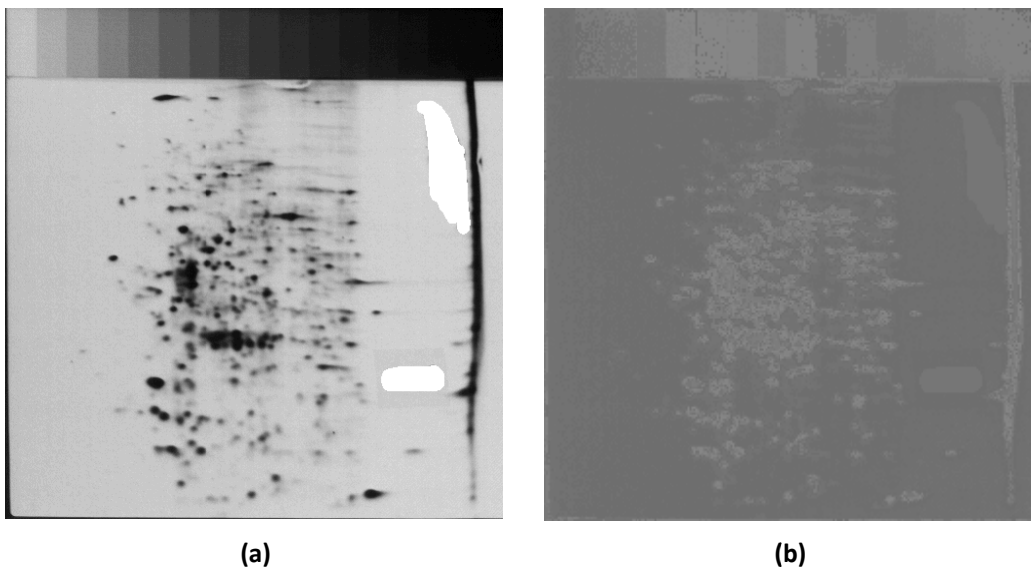
Figure 8: 2-D gel electrophoresis image of the first sample of Patient- Human leukemias

(a) Original, (b) Gradient image, (c) Gradient image after applying FWB Segmentation algorithm _without denoising_ (PSNR =21.767db), and (d) Gradient image after FWB Segmentation algorithm _with denoising_ (PSNR =26.0292db)

However, the visual inspection is not enough to judge the quality of the images. So, we will use the E_{CW} evaluation error to evaluate the performance of the proposed algorithm the Fuzzy-Watershed Based algorithm (FWB) with and without denoising. We set the threshold TH = 10 and the normalization factor Z = 1000000 as in[18].

Table 3: The intra region error (Eintra) of the Fuzzy watershed algorithm with and without denoising on seven data samples

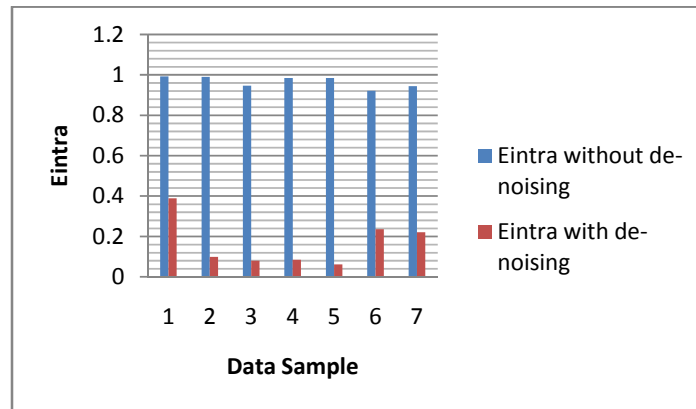| Data Sample no	Eintra (without denoising)	Eintra (with denoising)
1	0.9929	0.38911
2	0.98994	0.099243
3	0.94678	0.079956
4	0.98464	0.084473
5	0.98448	0.061584
6	0.92159	0.23535
7	0.94414	0.22076

Figure 7: The intra region error (Eintra) of the Fuzzy watershed algorithm with and without denoising on seven data samples

Table 4: The inter region error (Einter) of the Fuzzy watershed algorithm with and without denoising on seven data samples

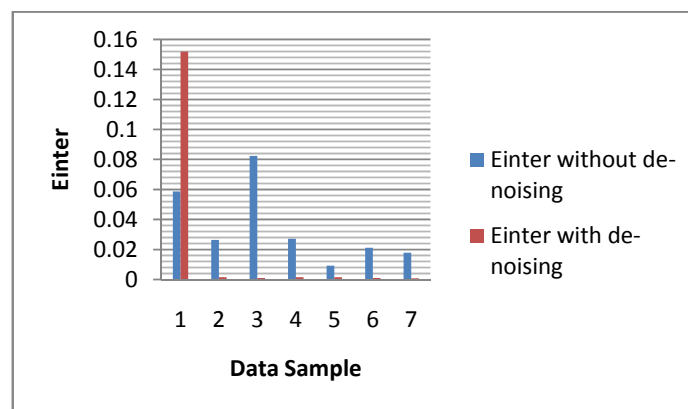| Data Sample no	Einter (without denoising)	Einter (with denoising)
1	0.05867	0.15187
2	0.02635	0.001455
3	0.08236	0.00102
4	0.02723	0.0014871
5	0.00921	0.0015185
6	0.02121	0.0010123
7	0.01788	0.00085073

Figure 8: The inter region error (Einter) of the Fuzzy watershed algorithm with and without denoising on seven data samples

From the previous two tables and figures, we can observe that performing the wavelet denoising step before the segmentation step improved the intra and the inter region error significantly in most of the used samples (only sample 1 produced worse **Einter**). The improvement in the **Eintra** (which evaluates the under-segmentation error) is high in all the seven cases, which means (100%) of the cases were improved with an improvement of about 9.3% as in cases (2,3,4 and 5) of the error value. This improvement means that approximately all spots appeared clearly in the resulted images.

The **Einter** (which evaluates the over-segmentation error) had been highly improved in six cases of the seven cases i.e the error was decreased with a (85.7%) improvement from all cases with an improvement of 8% as in case 3 of the error value.

We can also observe that in the first five data samples which are data for the human leukemia and human blood lymphocytes where exist the problems of ghost (weak) spots and noisy background, the application of denoising technique before the proposed algorithm, comparatively with the Fuzzy-Watershed algorithm without denoising, succeeded in reducing the problem of over-segmentation and identifying weak spots as in all data samples except for the first sample. For the last two data samples which are the Fetal Alcohol Syndrome where exist the problems of contamination of gels and overlapped spots, the application of denoising technique before the proposed algorithm, comparatively with the Fuzzy-Watershed algorithm without denoising, succeeded in reducing the problem of over-segmentation and identifying overlapped spots.

7. Discussion

In this work, we presented a new algorithm based on the notion of fuzzy relations to segment and detect protein spots in 2-D gel electrophoresis images. This algorithm shows high performance and detects the protein spots precisely. The new algorithm simplifies the original image to a mosaic image where applying the watershed algorithm, the number of catchment basins is reduced and hence the problem of over-segmentation is handled. Also, we can say that the addition of the denoising step yield to better results.

For future work, we suggest the development of fuzzy relations to obtain better results. The second relation can be a fuzzy relation defining the degree of membership of the grey value to a particular cluster for enhancement and improvement of the new algorithm.

References

[1] E.Bettens, P. Scheunders, J. Sijbers, D.Van Dyck, L.Moens." Automatic Segmentation and Modeling of twodimensional Electrophoresis gels ", in Proceedings of International Conference on Image Processing, 1996.

[2] Ming Hung Tsai et al., "Watershed-Based Protein Spot Detection in 2DGE images", in the proceeding of ICS 2006, International Workshop on Software Engineering, Databases, and Knowledge Discovery, Taiwan, 2006

[3] Luis Patino, "Fuzzy relations applied to minimize over

segmentation in watershed algorithms", Pattern Recognition Letters 26 (2005) 819–828

[4] Hartigan, J.A., 1975. "Clustering Algorithms". Wiley, New York.

[5] Kohonen, T., 1998. "The self-organizing map".Neurocomputing 21, 1–6.

[9] Valente de Oliveira, W. Pedrycz "Advances in Fuzzy Clustering and its Applications" published 2007

[10] Beucher, S., Lantuéjoul, C., "Use of watersheds in contour detection", Proc. Int. Workshop Image Processing, Real time edge and motion detection/estimation, Rennes, France, Sept. 17-21, 1979.

[11] Beucher, S., "Watersheds of functions and picture segmentation", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Paris, France, May 1982, pp.1928-1931.

[12] Minh-Tuan Trong Hoang, Yonggwan Won, "A Marker-Free Watershed Approach for 2D-GE Protein Spot Segmentation," isitc, pp.161-165, 2007 International Symposium on Information Technology Convergence (ISITC 2007), 2007

[13] D.T. Lin and J.L. Kuo, "Improved Watershed Algorithm Spot Detection on Protein 2D Gel Electrophoresis Images", Proceeding (444) Signal and Image Processing - 2004

[14] E.Bettens, P.Scheunders, J.Sijbres, D.Van Dyck, and L.Moens,"Automatic Segmentation and modeling of twodimensional Electrophoresis Gels," Proc. IEEE Int'l conf.on Image Processing, vol. 1, pp. 665-668, Sep . 16-19,1996.

[15] Ming-Hung Tsai, Hui-Huang Hsu, Chien-Chung Cheng, "Watershed-Based Protein Spot Detection in 2DGE images", in Proc. Int'l Computer Symposium (ICS 2006) , vol. III, p.p. 1334-1338, Taipei, Taiwan, Dec. 4-6, 2006.

[16] www.ccrnb.ncifcrf.gov/2DgelData Sets

[17] bioinformatics.org/lecb2dgeldb

[18] H.-C. Chen and S.-J. Wang, "The use of visible color difference in the quantitative evaluation of color image segmentation," in Proc. ICASSP, 2004.

[19] S. G. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. Inform.Theory, vol. 38, pp. 617–643, Mar. 1992.

[20] Stein, Charles M. (November 1981). "Estimation of the Mean of a Multivariate Normal Distribution". The Annals of Statistics 9 (6): 1135–1151

[21] Alessio Soggiu, Osvaldo Marullo, Paola Roncada and Enrico Capobianco," Empowering spot detection in 2DE images by wavelet denoising", In Silico Biology 9, 0011 (2009); Bioinformation Systems