International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **12**, December **2009**

 ISSN 1349-4198
pp. 1–ISII08-209

A SIMPLE METHOD FOR DETECTING PROTEIN SPOTS IN 2D-GE IMAGES USING IMAGE CONTRAST

Meng-Hsiun Tsai[1,2], Shu-Fen Chiou[3] and Min-Shiang Hwang[1,*]

[1]Department of Management Information Systems
[2]Institute of Bioinformatics
[3]Department of Computer Science and Engineering
National Chung Hsing University
250 Kuo Kuang Road, Taichung, 402, Taiwan
*Corresponding author: mshwang@nchu.edu.tw

Received July 2008; revised December 2008

Abstract. *The proteomics is a popular bioinformation research. In one specification state, one cell can be presented a kind of proteins, and we say that is a proteome. The proteome can be changed with different situations. In order to analyze the proteomics, the first step is that try to separate the proteins. The 2D-GE (Two-Dimensional Gel Electrophoresis) is an important and useful tool for research. Today, biologists have access to the many image databases that exist online. Biologists are keenly interested in the information that might be discerned from protein spots. Thus, quickly and efficiently finding such spots is an important job. In this paper, we attempt to find these spots based on 2D-GE images' features. We first use the image contrast to find the proteins' edges and then we detect the protein spots by the edges. Our experiments yield good results for detecting protein spots in 2D-GE images.*
Keywords: 2D-GE image, Two-dimensional gel electrophoresis, Protein spots.

1. **Introduction.** One popular technique utilized in biotechnology research is two-dimensional gel electrophoresis (2D-GE) [2, 4, 6, 7, 12, 15]. Researchers use 2D-GE to identify pathogenic proteins, as well as to aid in discovering cures for diseases caused by such proteins.

The 2D-GE utilizes the isoelectric point (PI) of iso-electric focusing (IEF) in the horizontal dimension and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in the vertical dimension. Using a scanner to scan the 2D-GE, a 2D-GE image can be obtained. First we separate the proteins' sports from 2D-GE images by computer's computation. We can find the interesting spots by comparison gels. Then we obtain the protein information such as protein mass, peptide sequence tag, etc. [9, 17]. Finally, we search the database and know that what is the protein with this spot. For any one disease, we compare 'normal' images with those from patients suffering from a disease. Biologists attempt to find abnormal proteins on the patient's 2D-GE images and compare them to normal ones. Once divergences are identified, a cure can be attempted.

The most important information in a 2D-GE image are the protein spots. Finding these protein spots on a 2D-GE image is thus a critical issue. There are some methods for this job [3, 8, 11]. In Bettens et al.'s [3] method, they use the Watershed method to detect the protein spots in 2D-GE images. In Persson and Bigun's method [11], they use the symmetry features to detect the spots. They are only based on general image processing methods to detect protein spots. They did not use the 2D-GE's features to their scheme. In this paper, we consider the salient features of 2D-GE images and use these features to

locate the protein spots. We use the contrast of the image to separate the protein spots and backgrounds. Then we detect and elucidate the spots.

This paper is organized as follows: Section 2 introduces 2D-GE and image contrast. We propose our method in Section 3. Section 4 gives the experimental results and discussion. The final section contains our conclusions.

2. Related Work.

2.1. 2D-GE images.

2D-GE [9, 17] is based on Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). There are two steps to generate 2D-GE. The first step is the first dimension generation. The proteins focus their iso-electric point (PI). The second step is the second dimension generation. This step separates proteins by their molecular weights. After obtaining the 2D-GE, we can use scanners or cameras to scan or picture the 2D-GE to get 2D-GE images. The Figure 1 [13] is an 8bits of gray level 2D-GE image. There are three elements in 2D-GE images: protein spots, background, and noise.

Protein spots: These are most important information contained in 2D-GE images. A protein spot in an image consists of an approximate dark circle. It is not monochromatic and homogenous. From center of the spot to edge, the intensity of its color decreases. Figure 2 is a protein spot. Every spot in Figure 1 is a kind of protein.

Noises: 2D-GE images often contain noise. Noise can result from winding gels, error scans, etc. Sometimes the noise is easily mistaken for protein spots. It is important but difficult to distinguish protein spots from noise.

Background: Aside from protein spots and noise, there is the background. The backgrounds' gray levels are very similar, and they are lighter than protein spots.

2.2. The watershed method.

The watershed method called the watershed transformation (WST) [16] is a segmentation algorithm for images. This algorithm usually can get the good results for segmenting the objects in the images. In 1997, Bettens et al. [3] use this concept to detect the protein spots in 2D-GE images.

2.3. The symmetry features.

In Persson and Bigun's method [11], they find the symmetry features with the 2D-GE images and use these features to detect the protein spots. For each pixel, they generate the feature vector which consists of the local LoG transform and local symmetry derivative response. Then they classification these vector to achieve the propose of detect protein spots.

2.4. Image contrast.

[14] proposed six features of texture images: coarseness, contrast, directionality, linelikeness, regularity, and roughness. Almost researchers use these features for image retrieval or other image processing[1, 5, 10, 19], but we think they can be used in other domains. Especially in the case of image segmentation, we think that contrast can be used for finding protein spots.

Comparing the colors in image, we can obtain degrees of chromatic differentiation, which we call the degree contrast. Every pixel in the image has its own contrast, and we can determine this by comparing them with the neighboring pixels. The contrast is defined in Equation 1.

$$F_{con} = \frac{\sigma}{(u_4/\sigma^4)^{\frac{1}{4}}} \tag{1}$$

In Equation 1, σ is the pixel values' standard error, and u_4 is fourth moment of the image.

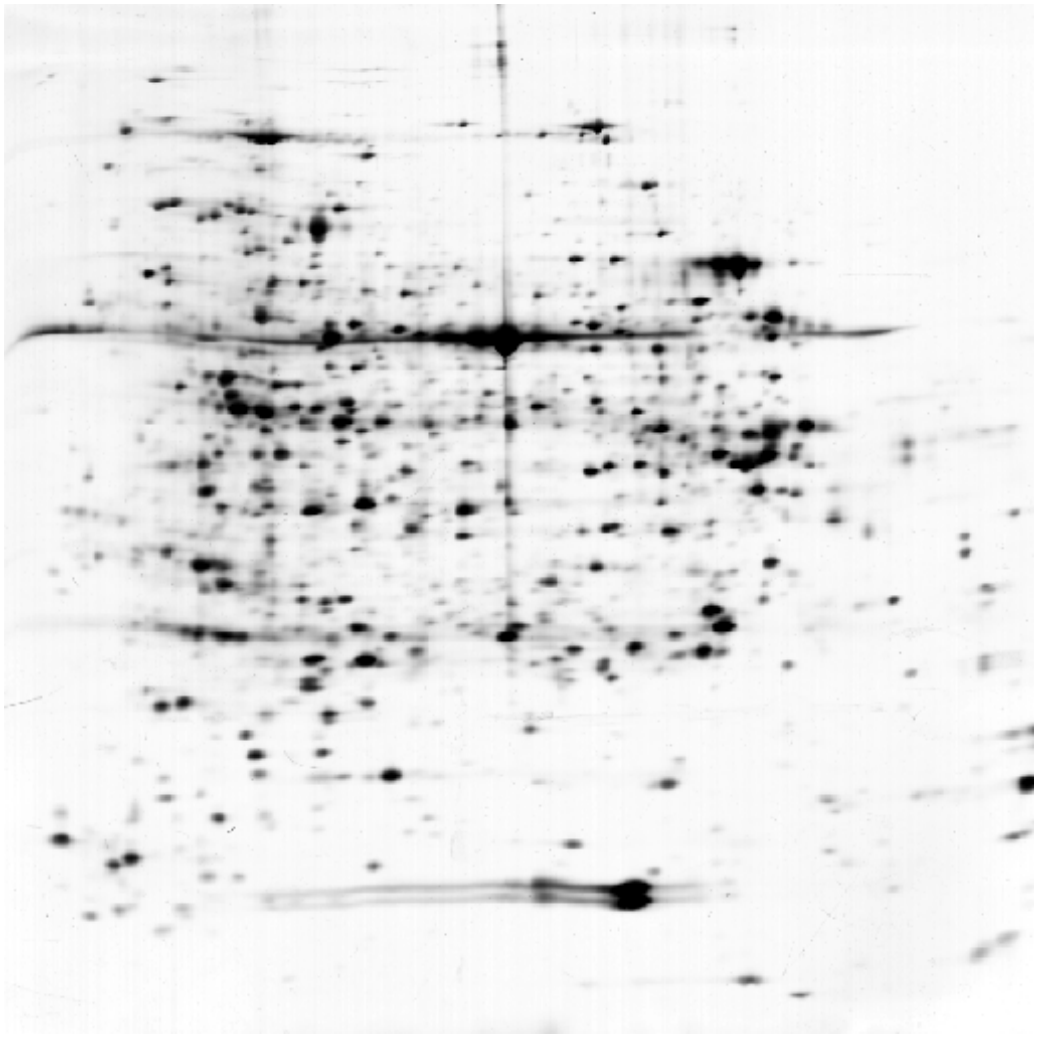


FIGURE 1. A 512 × 512 2D-GE image

FIGURE 2. A protein spot

3. **Proposed Method.** In this section we describe our method for detecting protein spots in 2D-GE images. In this paper, the N is the width of the image, and M is the image's height. $F_{con}(i, j)$ is contrast of the pixel in the ith row and jth column. Our method uses the following steps:

1. Computing the contrast.
2. Detection.

3.1. **Computing the contrast.** Considering each pixel to be a center, we compute the contrast among its eight neighboring pixels using Equation 1. After we compute all pixels in the image in this fashion, we determine the maximum and minimum contrasts. Then we normalize all contrasts by Equation 2. The max is the maximum contrast, and the min is the minimum contrast.

$$\frac{F_{con}(i, j)}{max - min} \tag{2}$$

Then we compute the average normalization contrast as the default threshold t, and take those pixels with contrasts greater than t as candidates to be protein spots.

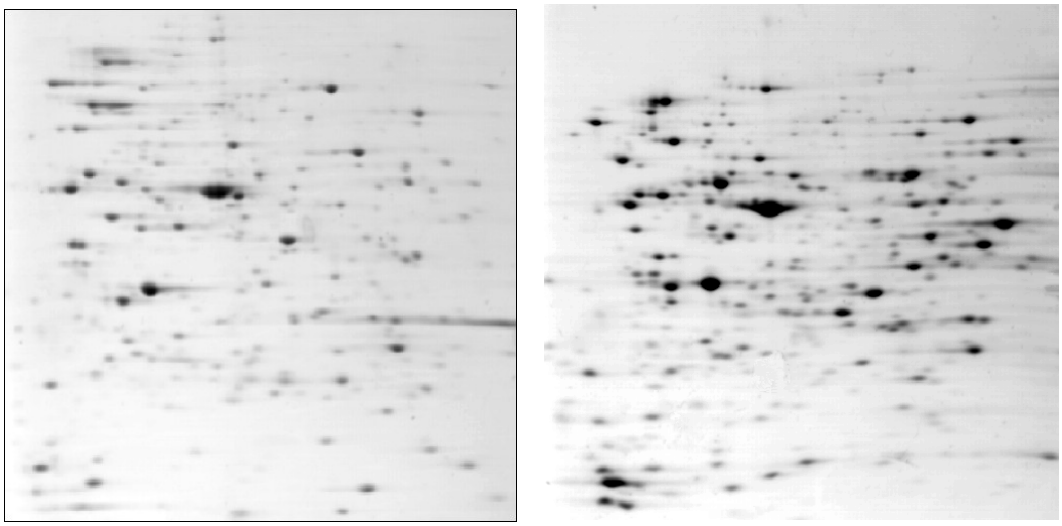
3.2. **Detection.** Once we identify those pixels whose contrast is greater than t, we determine the protein spots using the Algorithm 1:

Algorithm 1 Detecting protein spots

Input Pixels' contrasts k that greater than t
Output Protein spots $P = \{p_m, p_{m+1}, p_{m+2}, \ldots, p_n\}$
for m=0,. . . ,n
  while $k > 0$ **do**
    Take a pixel s in k as the p_mth protein spot
    while every pixel $\in p_m$ hasn't been found its eight neighbor **do**
      $k = k - 1, s \in p_m$
      Find the s's eight neighbor b
      if b in k then
        $b \in p_m$ and $k = k - 1$
      end if
      $m = m + 1$
    end while
  end while
end for

The s'eight neighbors are its eight neighboring pixels. Algorithm 1 yields the number of protein spots and relevant information about them.

4. **Experimental Results and Discussions.** In the first experiment, we use three 512 $\times$ 512 original 2D-gel gray images in Figures 1, 3 [20]; namely g_1, g_2, and g_3. The gray image's pixel value ranges from 0 to 255. In the first step, we calculate the pixels' contrasts within image.

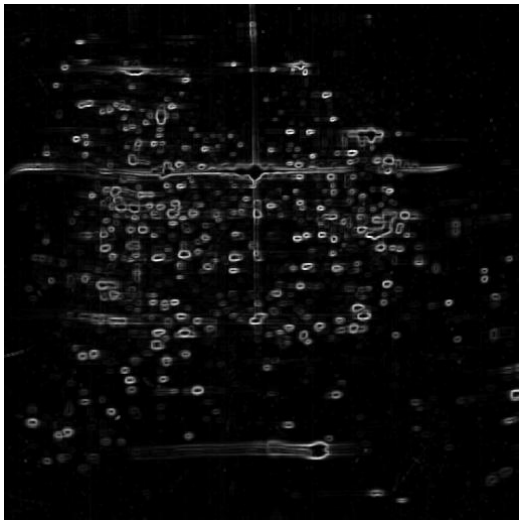


(a) g_2's original image (b) g_3's original image

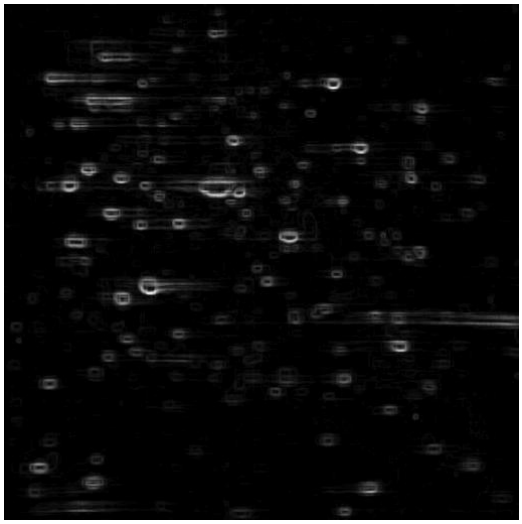
FIGURE 3. g_2 and g_3s' original images

We compute the pixel contrast from left to right, and top to bottom. Figures 4 (a), (b), and (c) are the g_1, g_2, and g_3s' display the contrasts of the images. The lighter pixels

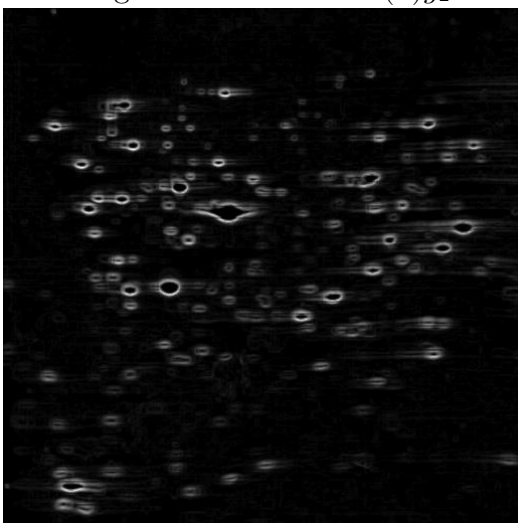
in the images are the relevant pixels. We then count them and compute the average threshold t.



(a) g_1's contrast image



(b)g_2's contrast image



(c) g_3's contrast image

FIGURE 4. g_1 g_2, and g_3s' contrast images

In Table 1, we use $1t$, $2t$, and $3t$ as threshold by which the pixels are measured to detect the protein spots in g_1, g_2, and g_3. The $1t$, $2t$, and $3t$ are 1 time, 2 times, and 3 times greater than the average contrast, respectively. We define one protein spot's range to be not greater than 500 pixels. Different thresholds in same image can detect different numbers of protein spots.

TABLE 1. multiple thresholds and number of protein spots

	$1t$	$2t$	$3t$
g_1	472	358	305
g_2	236	120	69
g_3	516	243	206

Figure 5 shows the g_3 with t =$1t$. As can be seen, six colors present the protein spots sequentially. Our goal is to detect more protein spots in the 2D-GE images, but we cannot

absolutely identify them. This is the job of a biologist. Some spots we detect may seem to be noise; but because we think that they may contain relevant information, we do not delete them.

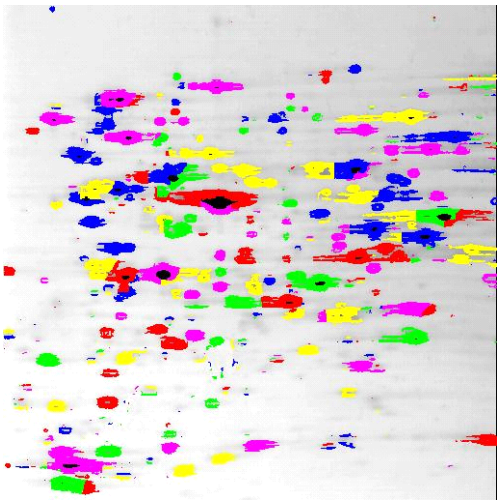


FIGURE 5. g_3 with $t = 1t$

In the second experiment, we get the 2D-GE images from the web "World-2DPAGE" [18]. The two 2D-GE images are cytoplasmic protein extracting from cells of corynebacterium glutamicum. And the PH values are PH 4.5-5.5 and PH 5-6. The images size are 1606×2008 pixels. These are also gray level (8-bits) images. Figure 6 is the original images. We also show their contrast images in Figure 7 with $t = 2t$.

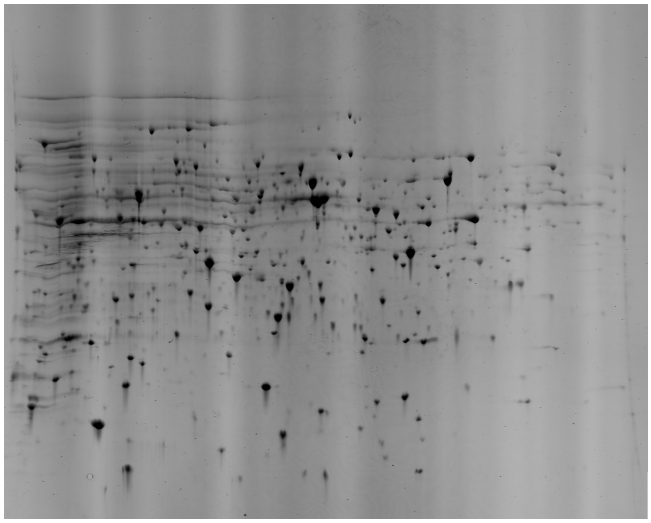
Figure 8 shows that the proteins which we detect using our method with $t = 2t$. We use five colors to draw the protein spots. In the Figure 8(a), the result is better than Figure 8(b), because the spots in Figure 8(b) look like the noises. In order to delete these possible noises in Figure 8(b), we calculate average area of detected protein spots as $avg - spot$. If the area of detected spots is equal or lower than 50% $ave - spot$, we consider this spot is a noise. After this process, we get the Figure 9 and the number of protein spots is 233.

TABLE 2. Number of corynebacterium glutamicum protein spots with $t = 2t$ and $t = 3t$

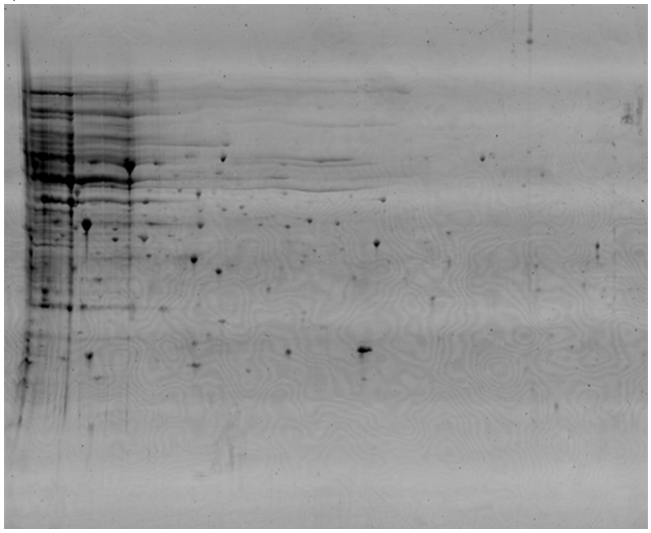
Corynebacterium glutamicum	$2t$	$3t$
$PH4.5 - 5.5$	784	618
$PH5 - 6$	374	164

In our experimental results, we can see the almost possible protein spots in 2D-GE by our eyes. When we mark the possible spots, we also need the biologist to ensure whether the possible spots are protein spots.

5. **Conclusions.** In this paper, we propose a simple method to detect protein spots in 2D-GE images based on the image's chromatic features. We use contrast to separate the protein spots and background. Then we compare contrast among pixels with a threshold. If their contrast is greater than threshold, they are candidate pixels. Finally we locate every candidate pixel among its eight neighbors and use these groupings of pixels to locate protein spots. Our method yields good results that can be evaluated even with the naked eye.



(a) Corynebacterium glutamicum with PH 4.5-5.5



(b) Corynebacterium glutamicum with PH 5-6

FIGURE 6. Original 2D-GE images with cytoplasmic protein extractin form corynebacterium glutamicum

REFERENCES

[1] P. K. Amin, N, Liu, and K, P. Subbalakshmi, Statistical attack resilient data hiding, *International Journal of Network Security*, vol.5, no.1, pp.112-120, 2007.

[2] G. Angelica, Advances in 2d gel techniques, *Proteomics: A Trends Guide*, pp.1471-1931, 2000.

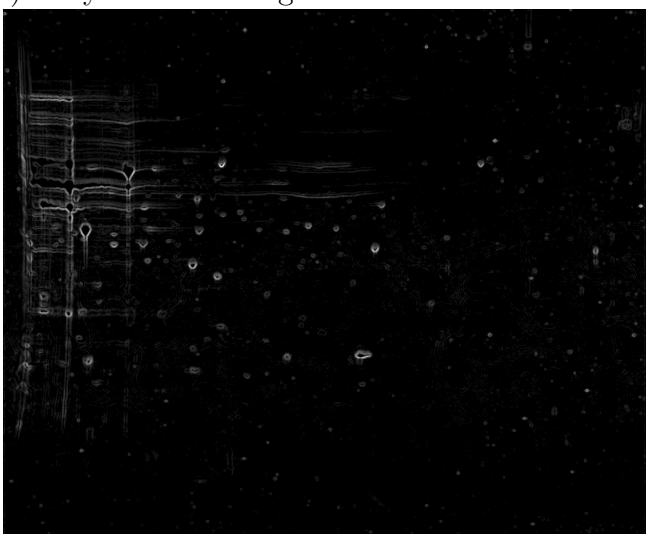
[3] E. Bettens, P. Scheunders, D. V. Dyck, L. Moens, and P. V. Osta, Computer analysis of two dimensional electrophoresis gels: A new segmentation and modelling algorithm, *Eletrophoresis*, vol.18, no.5, pp.792-798, 1997.

[4] T. S. Chen, P. I. Chen, and M. C. Hsiao, Automatic high dynamic range and resolution imaging for 2dge images, *Proc. of the IEEE 6th World Congress on Intelligent Control and Automation*, pp.10507-10511, 2006.

[5] N. E. Fishawy and O. M. A. Zaid, Quality of encryption measurement of bitmap images with RC6, MRC6, and Rijndael block cipher algorithms, *International Journal of Network Security*, vol.5, no.3, pp. 241-251, 2007.

[6] K. Kaczmarek, B. Walczak, S. d. Jong, and B. G. M. Vandeginste, Matching 2d gel electrophoresis images, *J. Chem. Inf. Comput. Sci.*, vol.43, pp.978-986, 2003.

(a) Corynebacterium glutamicum with PH 4.5-5.5



(b) Corynebacterium glutamicum with PH 5-6

FIGURE 7. Contrast 2D-GE images with cytoplasmic protein extractin form corynebacterium glutamicum, $t = 2t$

[7] N. Khan and S. Rahman, A new approach to detect similar proteins from 2d gel electrophoresis images, *Proceedings of Third IEEE Symposium on Bioinformatics and Bioengineering*, pp.182-189, 2003.

[8] Y. Kim, J. Kim, Y. Won, and Y. In, Segmentation of protein spots in 2D gel electrophoresis images with watersheds using hierarchical threshold, ISCIS 2003, LNCS 2869, pp.389-396, Springer-Verlag, 2003.

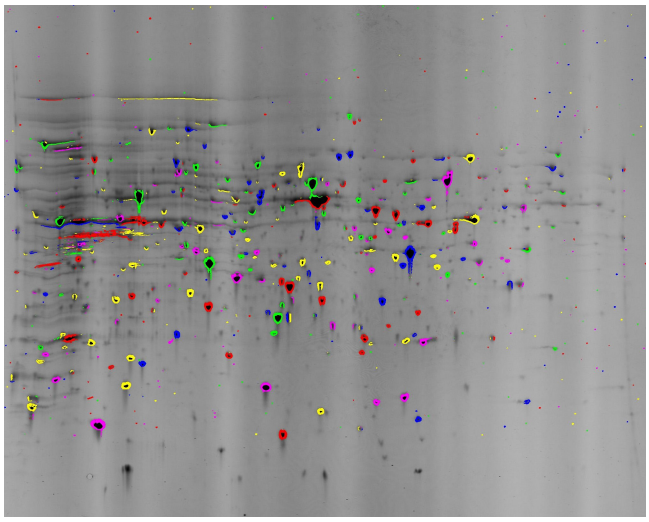
[9] B. Kuster and M. Mann, Identifying proteins and post-translational modification by mass spectrometry, *Current Opinions in Structual Biology*, vol.8, pp.393-400, 1998.

[10] X. Luo, M. Shishibori, F. Ren, and K. Kita, Incorporate feature space transformation to content-based image retrieval with relevance feedback, *International Journal of Innovative Computing, Information and Control*, vol.3, no.5, pp.1237-1250, 2007.

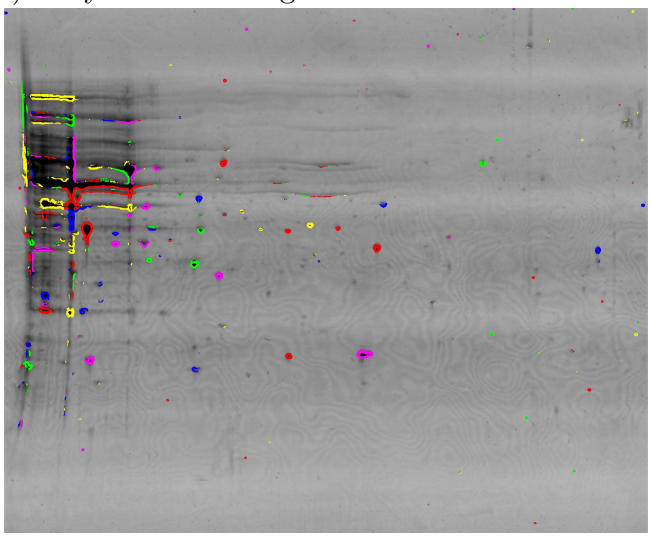
[11] M. Persson and J. Bigun, Detection of spots in 2-d electrophoresis gels by symmetry features, *Pattern Recognition and Data Mining*, LNCS 3686, pp.436-445, 2005.

[12] R. Stanislaus, C. Chen, J. Franklin, J. Arthur, and J. S. Almeida, Agml central: Web based gel proteomic infrastructure, *BIOINFORMATICS*, vol.21, no.9, pp.1754-1757, 2005.

[13] SWISS-2DPAGE. http://www.expasy.org/ch2d/



(a) Corynebacterium glutamicum with PH 4.5-5.5



(b) Corynebacterium glutamicum with PH 5-6

FIGURE 8. Detect protein spots with threshold $t = 2t$

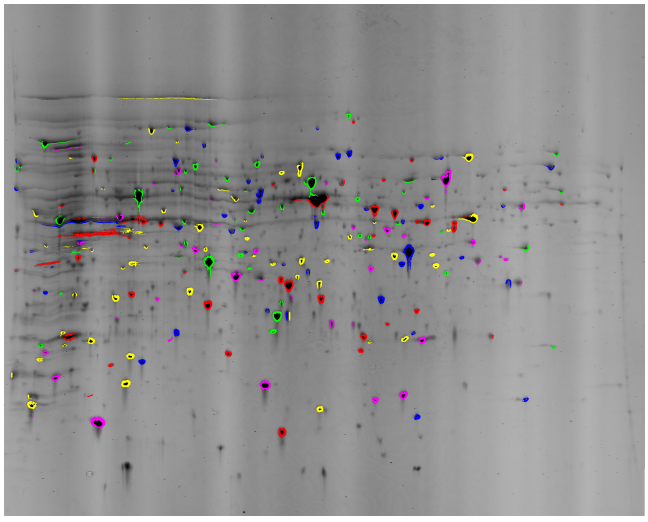


FIGURE 9. After deleting the possible noises in Figure 8(b)

[14] H. Tamura, S. Mori, and T. Yamawaki, Texture features corresponding to visual perception, *IEEE Transaction on Systems Man Cybernet*, vol.8, no.6, pp.460-473, 1978.
[15] M. H. Tsai, S. F. Chiou, and M. S. Hwang, A progressive image transmission Method for 2D-GE image based on context feature with different thresholds, *International Journal of Innovative Computing, Information and Control*, vol.5, no.2, 2009.
[16] L. Vincent and P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol.13, no.6, pp.583-598, 1991.
[17] M. R. Wilkins, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, *Proteome Research: New Frontiers in Functional Genomics*, Spriger-Verlag, 1997.
[18] World-2DPAGE. http://world-2dpage.expasy.org/repository/.
[19] P. Zhang, S. Hirai and K. Endo, A feature matching-based approach to deformation fields measurement from MR images of non-rigid object, *International Journal of Innovative Computing, Information and Control*, vol.4, no.7, pp.1607-1615, 2008.
[20] http://www.abdn.ac.uk/ mmb023/2dhaem/2dbkg_f.htm