Bayesian Inference for 2D Gel Electrophoresis Image Analysis

Ji Won Yoon[1], Simon J. Godsill[1], ChulHun Kang[2], and Tae-Seong Kim[3]

[1] Signal Processing Group, Engineering Department, Cambridge University, UK
[2] Graduate School of East-West Medical Science, Kyung Hee University, Korea
[3] Biomedical Engineering Department, Kyung Hee University, Korea

Abstract. Two-dimensional gel electrophoresis (2DGE) is a technique to separate individual proteins in biological samples. The 2DGE technique results in gel images where proteins appear as dark spots on a white background. However, the analysis and inference of these images get complicated due to 1) contamination of gels, 2) superposition of proteins, 3) noisy background, and 4) weak protein spots. Therefore there is a strong need for an automatic analysis technique that is fast, robust, objective, and automatic to find protein spots. In this paper, to find protein spots more accurately and reliably from gel images, we propose Reversible Jump Markov Chain Monte Carlo method (RJMCMC) to search for underlying spots which are assume to have Gaussian-distribution shape. Our statistical method identifies very weak spots, restores noisy spots, and separates mixed spots into several meaningful spots which are likely to be ignored and missed. Our proposed approach estimates the proper number, centre-position, width, and amplitude of the spots and has been successfully applied to the field of projection reconstruction NMR (PR-NMR) processing [15,16]. To obtain a 2DGE image, we peformed 2DGE on the purified mitochondiral protein of liver from an adult Sprague-Dawley rat.

1 Introduction

Recent advances in proteomics play a key role in life science by identifying and characterizing overall proteins, and provide insights of disease and drug interactions. 2DGE is a widely used technique to analyze the protein complexes in proteomics and bioinformatics [11,8]. The two dimensions in 2DGE correspond to isoelectric point and mass: the isoelectric point separates the proteins in terms of a gradient of pHs, and the mass according to the weights of proteins. 2DGE yields an image representing the distribution of protein spots.

The 2DGE image analysis includes spot detection, segmentation, characterization, quantification, and etc. However, such analysis are complicated for the following reasons. First, there may be weak and small spots which are not be detected. Second, spots can be superimposed. These mixed spots are hard to separate by inspection or many deterministic approaches and the mixed spots are often likely to be regarded as one big spot. Even though interesting spots are clearly visible, it is difficult to recognize them if they are mixed with other

S. Hochreiter and R. Wagner (Eds.): BIRD 2007, LNBI 4414, pp. 343–356, 2007.

spots. Finally, there are spots not discernible from background noise. Spots in 2DGE image may be corrupted by two kinds of noise: global noise and local noise. Global noise is a background noise which has a specific pattern. Local noise effects the intensity of a pixel or a small area of image. Thus, we may have to restore images and find important spots with careful consideration of noise.

Many researchers have been worked on 2DGE image processing and analysis using several methods such as filtering in the wavelet domain [12], watershed techniques [14] and pixel value collection [10]. However, they are not good enough in noisy images and produce only limited results such as segmentation or quantification. In this paper, we tackle these problems by applying Reversible Jump Markov Chain Monte Carlo (RJMCMC) [2,7]. This method has been successfully applied to Projection-reconstruction NMR (PR-NMR) to reconstruct NMR spots in 2D signals [15,16]. Our application of RJMCMC to 2DGE attempts to subtract background noise from the image to enhance weak spots. That is, RJMCMC searches for weak spots which are likely to be ignored due to their weakness. The method also finds the proper number of spots automatically, restores the noisy images, and unmixes spots into more meaningful ones. In modeling of spots, we assumed Gaussian shape, as applied in other studies [13,9]. The assumption of such a specific shape of a spot has a significant benefit in that, since it is robust against local noise, we can estimate the signals based on interesting areas rather than each pixel. In our proposed RJMCMC method, we have incorporated the following characteristics to meet the requirements of gel image analysis: dimension invariant and moves in birth, death, split, and merge.

This paper consists of three main parts. In the first part, we present the mathematical model for 2D gel electrophoresis. Next, we describe the main algorithms for RJMCMC. Finally, the synthetic and experimental results from RJMCMC are demonstrated.

2 Model for 2D Gel Electrophoresis

As shown in Fig. 3, a typical 2D gel image contains numerous protein spots which might be individual or mixed. Therefore in our study, we model the image as a mixture of spots with Gaussian profile as follows:

$$I(x) = \sum_{k=1}^{K} A_k \phi_k(x; \mu_k, \Sigma_k) + \epsilon_e(x) \tag{1}$$

$$\begin{cases} x = [x_1, x_2]^T \\ \mu_k = [\mu_{k,1}, \mu_{k,2}]^T \\ \phi_k(x; \mu_k, \Sigma_k) \\ = \frac{1}{\sqrt{2\pi\Sigma_k}} \exp\left\{-\frac{1}{2}(x_1 - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right\} \\ \Sigma_k = diag(\sigma_{k,1}^2, \sigma_{k,2}^2) \\ \epsilon_e(x) \sim N(\epsilon_e(x); \mu_e, \sigma_e^2) \end{cases}$$

where $I(x)$ is the intensity at position x of the image. A_k is the amplitude of each spot, and $\phi_k(x; \mu_k, \Sigma_k)$ denotes the radial functions with a specific shape such as Gaussian, Lorentzian, or Laplacian shape. In this paper, we use a Gaussian

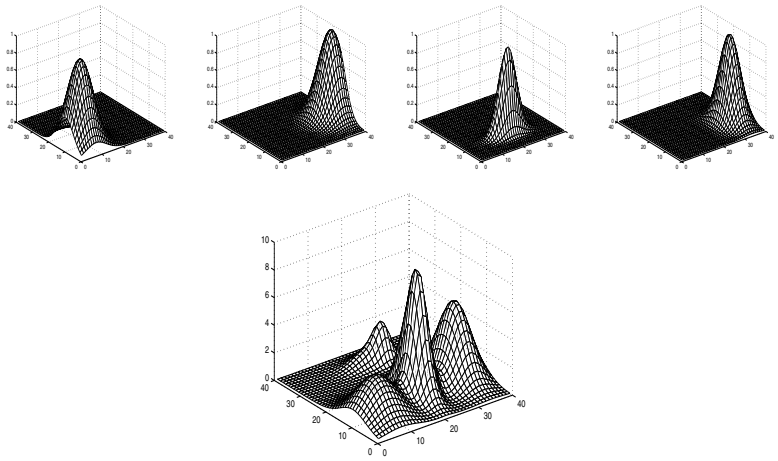

Fig. 1. Combination of radial functions: 4 different radial images with a spot are combined after multiplying their amplitudes - all images are vectorized in the linear model

shape for each spot. The radial function for each spot consists of two components, centre position μ_k and width of the spot σ_k. We assume that all spots have different spot widths and each spot has different spot width in terms of the x_1 and x_2 axes. $\epsilon_e(x)$ is a white noise, at position x, which is generated from normal distribution with a mean μ_e and a standard deviate σ_e. N denotes a normal distribution. RJMCMC estimates the μ_e and σ_e during its simulation automatically. Eq. (1) is well represented by Fig. 1.

In practice, Eq. (1) may be written in the linear model framework as follows:

$$Y = XA_{1:K} + \epsilon_e \text{ where } \epsilon_e \sim N(\epsilon_e; \mu_e, \sigma_e^2 I) \tag{2}$$

$A_{1:K} = [A_1, A_2, \cdots, A_K]$ is a vector for amplitudes of spots. X and Y are built from Eq. (1) by assembling all intensities $I(x)$ into a vector Y. That is, Y is a vector of noisy image intensities. X is defined by $[X_1, X_2, \cdots, X_K]^T$ where X_k is a vector made up from the image profile of spot k.

Denote by $\theta_k \in \Theta_k$, the parameter vector associated with the model indexed by $k \in \kappa$. Then, the priors are defined as

$$\begin{aligned} \theta_{1:K} &= (\mu_{1:K}, A_{1:K}, \sigma_{1:K}) \\ &\text{where } \begin{cases} \mu_{1:K} = (\mu_{1,1:K}, \mu_{2,1:K}) \\ \sigma_{1:K} = (\sigma_{1,1:K}, \sigma_{2,1:K}) \end{cases} \\ A_k &\overset{iid}{\sim} N(A_k; \mu_A, \sigma_A^2) \\ \mu_{1,k} &\overset{iid}{\sim} U(\mu_{1,k}; 0, T_1) \quad \mu_{2,k} \overset{iid}{\sim} U(\mu_{2,k}; 0, T_2) \\ \sigma_{1,k} &\overset{iid}{\sim} G(\sigma_{1,k}; \alpha, \beta) \quad \sigma_{2,k} \overset{iid}{\sim} G(\sigma_{2,k}; \alpha, \beta) \end{aligned} \tag{3}$$

where $k = 1, 2, \cdots, K$ and $K \in \{0, \cdots, K_{max}\}$. T_1 and T_2 are the size of an image. N, U and G stand for the normal, uniform, and gamma distributions respectively. α and β are assumed known. μ_A and σ_A are estimated during simulation.

Now, we may remove nuisance parameters $A_{1:K}$ by linear analytical integration since $A_{1:K}$ are assumed linear Gaussian,

$$P(\mu_{1:K}, \sigma_{1:K}|Y) = \int P(\mu_{1:K}, \sigma_{1:K}, A_{1:K}|Y) dA_{1:K} \tag{4}$$

The removal of the nuisance parameters makes RJMCMC more efficient (Rao-Blackwellization) [3] . The removed nuisance parameters $A_{1:K}$ are sampled from their full conditional $P(A_{1:K}|\mu_{1:K}, \sigma_{1:K}, Y)$, when required for estimation.

2.1 Likelihood

Marginalising the nuisance parameters, the likelihood is defined as follows:

$$\begin{aligned} P(Y|X, \mu_A, \Sigma_A, \Sigma_e) = & \frac{1}{(2\pi)^{(T_1 T_2)/2} |\Sigma_e|^{1/2} |\Sigma_A|^{1/2} |\Phi|^{1/2}} \\ & \times \exp\left\{ -\frac{1}{2} (Y'^T \Sigma_e^{-1} Y' + \mu_A^T \Sigma_A^{-1} \mu_A - \Phi^T \hat{A}) \right\} \\ & \text{where} \begin{cases} \hat{A} = \Phi^{-1} \phi \\ \Phi = X^T \Sigma_e^{-1} X + \Sigma_A^{-1} \\ \phi = X^T \Sigma_e^{-1} Y' + \Sigma_A^{-1} \mu_A \\ Y' = Y - \mu_e \end{cases} \end{aligned} \tag{5}$$

where $\Sigma_A = \sigma_A^2 I$ and $\Sigma_e = \sigma_e^2 I$ respectively.

3 Algorithms

Since one does not know the exact number of spots in a given image, the number must be estimated during the processing. That is, to calculate the proper number of spots is equivalent to estimate the exact dimension of the parameters as they are proportional to each other. This kind of problem is addressed in trans-dimensional approaches. One of the best known trans-dimensional approaches is a generalization of Markov Chain Monte Carlo method, so called Reversible Jump Markov Chain Monte Carlo method (RJMCMC) [2,7]. RJMCMC proposes a next state given by current state in the time series and it constraints on Markov chains. In this paper, RJMCMC has several moves to find parameters of interests and their dimensions: Birth, Death, Split, Merge, and invariant moves. The next image is proposed from the current image via these moves. The Birth move makes a new spot in the current image randomly in terms of a given proposal distribution. The Death move deletes an existing spot in the current image. Birth and Death moves are designed to satisfy reversibility conditions. The Split move divides a spot in the current iteration into two different spots in the next iteration. Conversely, Merge move makes two selected spots in the current iteration into a single spot in the next iteration. Split and Merge moves have reversibility conditions as for Birth and Death moves [2]. The last move is a dimension invariant move, i.e. it does not change the dimension of the

parameters. Instead, each parameter is sampled by a standard Markov Chain Monte Carlo (MCMC) step.

RJMCMC for 2D gel electrophoresis image has the following procedure in this paper:

- Propose a type of move from Birth, Death, Split, Merge, and Dimension invariant.
- If the move type is Dimension invariant, RJMCMC samples parameters using a standard Metropolis-Hastings (MH) algorithm, so that each unknown parameter is updated according to an acceptance probability.

$$\alpha_K = \min\left\{1, \frac{P(Y|\theta'_{1:K})P(\theta'_{1:K})q(\theta_{1:K};\theta'_{1:K})}{P(Y|\theta_{1:K})P(\theta_{1:K})q(\theta'_{1:K};\theta_{1:K})}\right\} \tag{6}$$

- If the move type is one of Birth, Death, Split, and Merge, RJMCMC follows a generalized MH step with an acceptance probability.

$$\alpha_{K'} = \min\left\{1, \frac{P(Y|K', \theta'_{1:K'})P(K')P(\theta'_{1:K'}|K')q_1(K;K')q_2(\theta_{1:K};\theta'_{1:K'})}{P(Y|K,\theta_{1:K})P(K)P(\theta_{1:K}|K)q_1(K';K)q_2(\theta'_{1:K'};\theta_{1:K})}\right\} \tag{7}$$

3.1 Dimension Invariant Move

RJMCMC is the same as standard MCMC in the case of Dimension invariant moves in that the dimension is fixed in Eq.(6). The Dimension invariant move samples two types of parameters, $\mu_{1:K}$ and $\sigma_{1:K}$. The prior structure of $\theta_{1:K}$ is assumed to be $P(\theta_{1:K}) = \prod_{k=1}^{K} P(\mu_k)P(\sigma_k)P(A_k)$, see Eq. (4) The kernel function $q(\theta'_{1:K};\theta_{1:K})$ proposes parameters using a Metropolis Hastings algorithm within the Gibbs method:

Dimension invariant Move

- Propose new parameters $\theta'_{1:K}$.
  - $\mu'_{1:K} \sim N(\mu'_{1:K}; \mu_{1:K}, \gamma^2 I)$
  - $\sigma'_{1:K} \sim G(\sigma'_{1:K}; \alpha, \beta)$ where α and β are known.
- Calculate likelihood from Eq. (5).
- Obtain α_K from Eq. (6).
- $u \sim U(u; 0, 1)$
- if $\alpha_K > u$, then
  - Accept the proposed parameters and replace them by the current parameters.
- else
  - Reject the proposed parameters and maintain the current parameters.

3.2 Other Moves: Birth, Death, Split, and Merge Moves

In Eq. (7), the prior distribution for the dimensionality, $P(K)$ is assumed uniform. From experience, convergence is rather slow when $\mu_{1:K}$ and $\sigma_{1:K}$ are updated jointly. Thus, only $\mu_{1:K}$ is sampled in the dimension variant moves (but σ_k is also proposed from the prior in the Birth move for a new spot) and $\sigma_{1:K}$ is updated in the dimension invariant move. These four moves assume the prior distribution $P(\mu_{1:K}|K)$ is a uniform distribution.

The kernel functions for dimension, $q_1(K^{'}|K)$, and parameters, $q_2(\theta^{'}_{1:K^{'}}|\theta_{1:K})$ are designed as follows in Birth and Death moves. The Birth move creates a new spot so that $q(K^{'} = K + 1|K) = 1$. The Death move however selects a spot randomly to delete among the K existing spots in the current step. Hence, the proposal probability of dimension for the death move is $q(K^{'} = K-1|K) = 1/K$. In the Birth move, the new spot $K^{'}$ is proposed from the prior, i.e. $q(\mu^{'}_k, \sigma^{'}_k) = p(\mu^{'}_k, \sigma^{'}_k)$.

Birth move

- $\mu_{K'}$ and $\sigma_{K'}$ are proposed for spot position and width.
- Calculate likelihood from Eq. (5).
- Obtain $\alpha_{K'}$ in Eq. (7).
- Let $u \sim U(u; 0, 1)$
- if $\alpha_{K'} > u$, then
  - Accept the proposed parameters and replace them by the current parameters.
- else
  - Reject the proposed parameters and maintain the current parameters.

Death move

- Select one among K spots and remove it.
- Calculate likelihood from Eq. (5).
- Obtain $\alpha_{K'}$ in Eq. (7).
- $u \sim U(u; 0, 1)$
- if $\alpha_{K'} > u$, then
  - Accept the proposed parameters and replace them by the current parameters.
- else
  - Reject the proposed parameters and maintain the current parameters.

The Split and Merge moves are related in a similar way. The first transition kernel function $q_1(K^{'}|K)$ is defined to be $1/K$ for both moves. The Split move divides a single spot, randomly chosen from the K existing spots, into two spots. In the Merge move, a single spot is randomly chosen and merged with its closest

neighbours. The Split kernel for $q_2(\theta^{'}_{1:K'}|\theta_{1:K})$ divides spot k into spots k and m, as follows:

$$\mu^{'}_m \sim q_2(\mu^{'}_m|\mu_k) = N(\mu^{'}_m; \mu_k, \lambda)$$
$$\mu^{'}_k \sim q_2(\mu^{'}_k|\mu_k) = N(\mu^{'}_k; \mu_k, \lambda)$$

where λ is assumed known. The Merge Kernel for $q_2(\theta^{'}_{1:K'}|\theta_{1:K})$ combines a spot k with its the nearest neighbour m, as follows:

$$\mu^{'}_k \sim q_2(\mu^{'}_k|\mu_m, \mu_k) = N(\mu^{'}_k; \bar{\mu}, \nu)$$
$$\text{where } \begin{cases} \bar{\mu} = \mu_m \times \omega_m + \mu_k \times \omega_k \\ \omega_m = \frac{A_m}{A_m + A_k},\ \omega_k = \frac{A_k}{A_m + A_k} \end{cases} \tag{8}$$

where ν is assumed known. In both moves, σ_k parameters are proposed from the prior.

Split move

- Select one among K spots and divide it into two spots in Eq. (8)
- Calculate likelihood from Eq. (5).
- Obtain $\alpha_{K'}$ as in Eq. (7).
- $u \sim U(u; 0, 1)$
- if $\alpha_{K'} > u$, then
  - Accept the proposed parameters and replace them by the current parameters.
- else
  - Reject the proposed parameters and maintain the current parameters.

Merge move

- Select one among K spots and search for its closest neighbour.
- Merge the two selected spots from Eq. (8)
- Calculate likelihood from Eq. (5).
- Obtain $\alpha_{K'}$ in Eq. (7).
- $u \sim U(u; 0, 1)$
- if $\alpha_{K'} > u$, then
  - Accept the proposed parameters and replace them by the current parameters.
- else
  - Reject the proposed parameters and maintain the current parameters.

Owing to the non-unique labeling of individual spots in the model Eq. (1), it is likely that spots become re-ordered during sampling, especially in a RJMCMC procedure where spots can be detected or added at each iteration. In order to

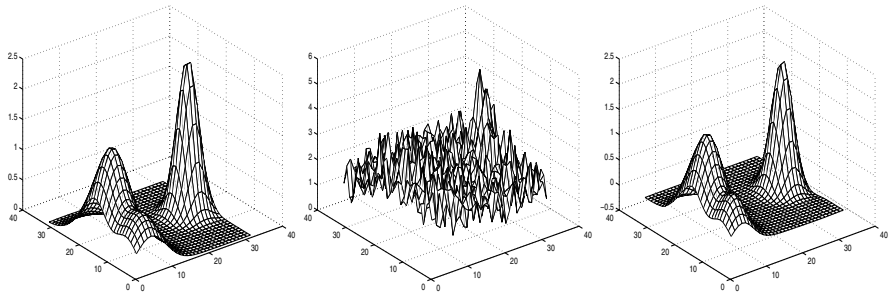

Fig. 2. A synthetic image with 4 spots : a pure image (a), a noisy image (b), an estimated image by RJMCMC (c)

address this labeling problem, we run a fixed dimensional RJMCMC with invariant moves after variant dimensional RJMCMC. That is, the variant dimensional RJMCMC generates the number of spots and the initial parameters for the fixed dimensional RJMCMC (see e.g. [5] for a detailed theoretical treatment of such issues).

4 Results

4.1 Synthetic Data

We tested RJMCMC performance on a synthetic image in Fig. 2. The first synthetic image has 4 spots, one very large spot and three overlapping spots. All spots have a Gaussian shape defined by centre position, width, and amplitude of the spot. The size of this image is 32 by 32. White Gaussian noise is added to the pure image with mean 2 and standard deviation 0.5. 2000 iterations are performed to address this problem including 1000 burn-in and 1000 for estimation. The estimated mean of noise is 2.0364. The left, centre and right figures denote a pure image without noise, corrupted image with noise generated by $\mu = 2$ and $\sigma = 0.5$ and a restored image from RJMCMC. The error is calculated by

$$\epsilon = ||\hat{S} - S||, \tag{9}$$

where $\hat{S}$ and S are the estimated image from RJMCMC and the original image without noise. We obtain $\epsilon = 1.8422$ after simulation for the first synthetic data set.

4.2 Experimental Data

To obtain experimental 2DGE images, we performed 2DGE on the purified mitochondiral protein of liver from an adult Sprague-Dawley rat. Fig. 3 shows that the resultant GE images present numerous protein spots.

We used two different concentrations of the sample and the images (a) and (b) of the figure show the results of $100\mu g$ and $200\mu g$ respectively. As we can

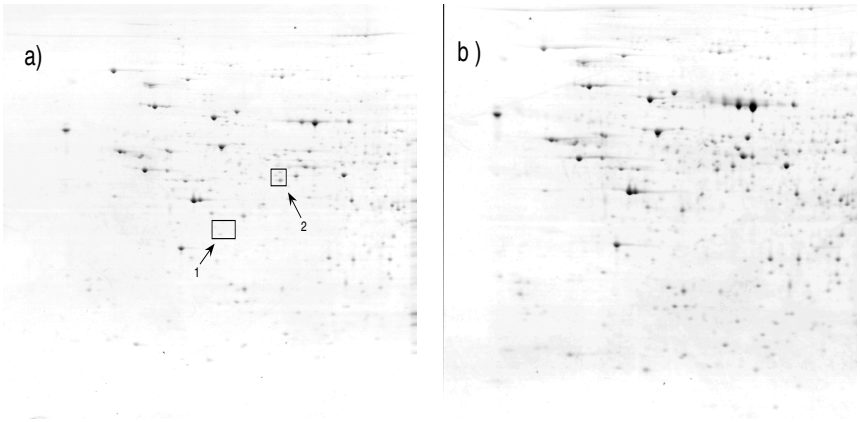

Fig. 3. Two gel images with a protein sample of $100\mu g$ (a) and $200\mu g$ (b): two rectangles (1 and 2) of the left image are used to search for the underlying spots such as weak spots and mixed spots in a mixed shape spot using RJMCMC

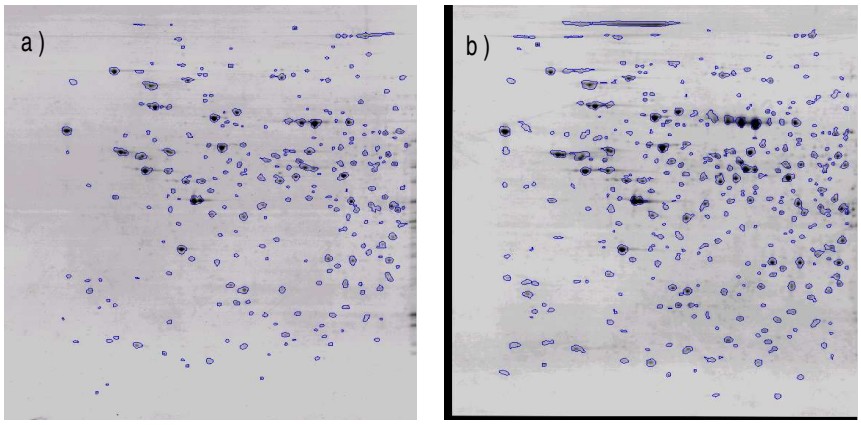

Fig. 4. Detection by commercial software (*Imagemaster 2D elite software v3.1*) for two gel images with a protein sample of $100\mu g$ (a) and $200\mu g$ (b)

see, the GE image (b) reveals more spots and much clearer than the image (a) due to the higher concentration which we intend to use in our evaluation and validation of the proposed methodology for detecting and identifying the spots.

To make a comparison against our proposed technique, the GE images in Fig. 3 were analyzed using a commercial gel image analysis software, Imagemaster 2D elite software v3.1 [6]. The images in Fig. 4 show the detected spots from each image. It is clear that there are more spots detected in (b).

For the RJMCMC image analysis, we selected the identical two sub-images from Fig. 3 (a) with the lower concentration as indicated in the figure with boxes. Before applying the RJMCMC, to remove the local variations or noise in

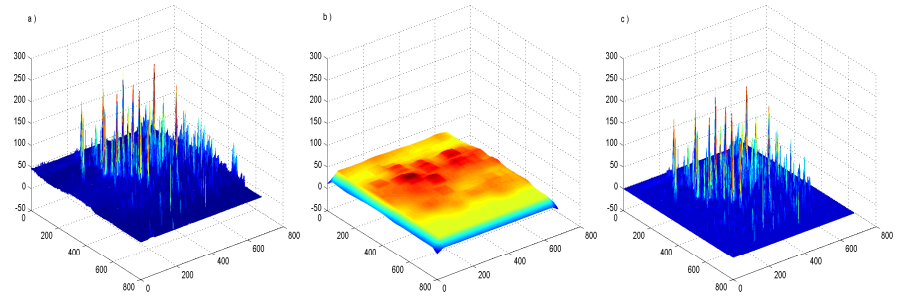

Fig. 5. Subtracting background noises: raw image (a), subtracted background noise (b), and extracted image (c)

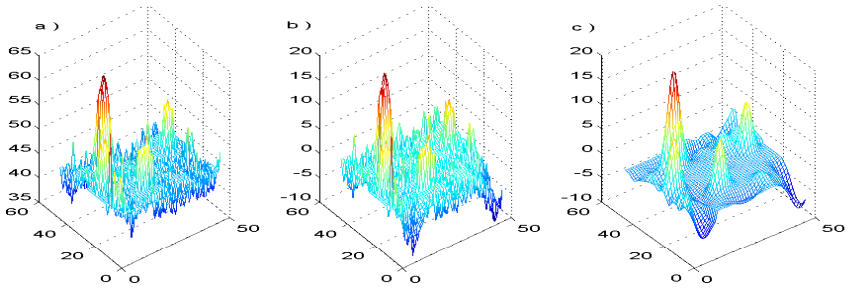

Fig. 6. Three images for the first example: a raw image (a), a filtered image without background noise (b) and a sample image in RJMCMC run (c)

the gel image, which creates a discrepancy between the real GE image and the theoretical model in Eq. (1), we used a simple approach, *local mean removal*, in which the average pixel intensity in local areas is removed [1]. Fig. 5 shows an example of local background noise removal. The RJMCMC was applied to the local noise-removed images with 10000 iterations including 5000 sampling and 5000 burn-in for sub-images. After the RJMCMC simulation, we run the fixed dimensional MCMC with 10000 iterations including 5000 sampling and 5000 burn-in for the labeling of the spots.

Fig. 6 and 8 show the steps of our RJMCMC algorithm. The image (a)s are selected from original 2DGE and the image (b)s are resulant images after removing background noise. The image (c)s are averaged images of samples from RJMCMC analysis. One can clearly notice that the RJMCMC generated images reveal potential spots with much better clarity. Furthermore, the RJMCMC generates statistical inferences to these spots.

Fig. 7 and 9 show the comparisons of the RJMCMC-inferenced spots against the detected spots by the commercial software. We used threshold to plot circles for the spots over 2.5 and 5 of the amplitudes (intensities) for Fig. 7 and 9 respectively.

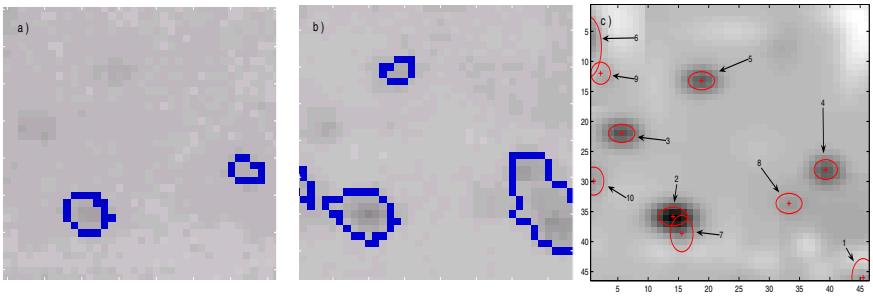

Fig. 7. comparison of detection for the first example: detection by commercial software for $100\mu g$ (a), detection by commercial software for $200\mu g$ (b) and detection by RJMCMC method for $100\mu g$ (c)

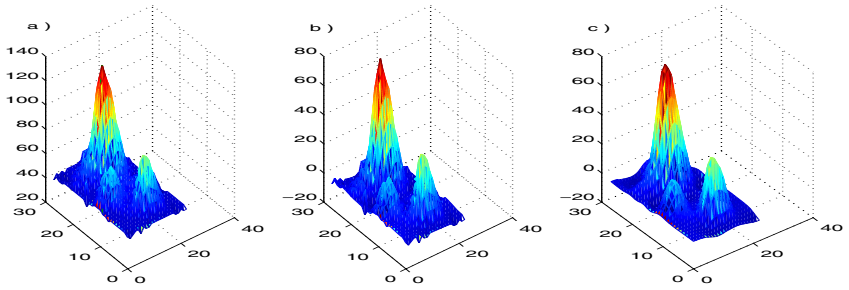

Fig. 8. Three images for the second example: a raw image (a), a filtered image without background noise (b) and a sample image in RJMCMC run (c)

Fig. 7 demonstrates that the RJMCMC can detect even very weak spots which would be unlikely to be detected in the conventional approaches. Note that the RJMCMC was applied to the gel image with a protein sample of $100\mu g$. The RJMCMC inferences numerous spots, as shown in Fig. 7 (c), that are not detected by the conventional software as in Fig. 7 (a). For instance Spot no. 5 and 10 are detected by the RJMCMC, but they are not detected in the (a), the lower concentration GE image. However, the higher concentration GE image confirms the presence of the spot in Fig. 7 (b). For Spot no. 2 and 7, the same spot has been detected in all three images, but the RJMCMC indicates the spots are composed of two spots. Additionally the RJMCMC indicates there could be more spots which are not visible in (a) and (b). It is not clear at the moment whether those RJMCMC inferenced weak spots are real or artificial, but it is clear that RJMCMC provides much higher sensitivity toward the detection of possible protein spots.

The results in Fig. 9 show the analysis of the mixed or overlapped spot in the selected region. The conventional software cannot separate the mixed and cannot give information about the possible individual profiles in both the low

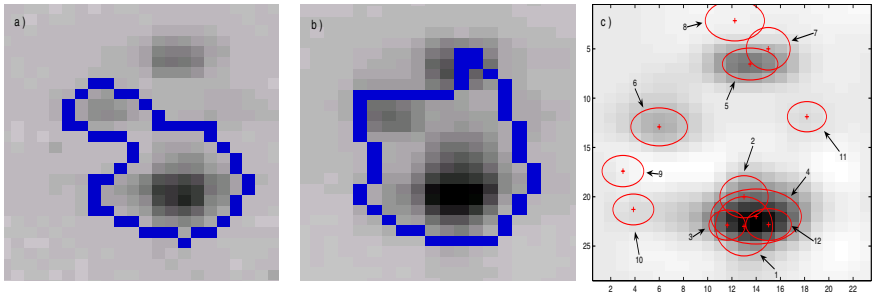

Fig. 9. comparison of detection for the second example: detection by commercial software for $100\mu g$ (a), detection by commercial software for $200\mu g$ (b) and detection by RJMCMC method for $100\mu g$ (c)

Table 1. Spot information for the first sub-image generated from RJMCMC

num	position (row)	position (col)	width (row)	width (col)	amplitude
1	45.5077	46.0000	1.8687	3.1683	3.5181
2	14.1045	35.8273	2.4262	1.5338	17.4910
3	5.6448	21.9955	2.1765	1.5216	10.8062
4	39.3266	28.0185	1.9230	1.6635	10.0152
5	18.8038	13.2272	2.1765	1.5122	9.8300
6	0	7.6177	2.2833	5.0953	15.3704
7	15.5598	38.6679	1.8768	3.0360	4.5647
8	33.2575	33.6518	2.1495	1.6844	2.6860
9	2.1577	11.9763	1.6069	1.8371	5.2277
10	1.0237	29.9370	1.6547	2.3283	3.3920

and high concentration GE images. Whereas the RJMCMC method may resolve each clustered spot into several individual spots as shown in (c).

Finally, our RJMCMC method generates the databases which are shown in Table. 1 and 2. Each table has six columns: index number, position for row, position for column, width for row, width for column and amplitude (intensity) of spots. As we can see in Tables, the amplitudes (intensities) of spots vary from 2.6 to 17.4 and from 5.7 to 503.9 for the Table. 1 and 2 respectively.

5 Conclusions

RJMCMC for 2DGE image analysis has two salient products: restoration and spot finding. 2DGE images suffer from high levels of noise yet RJMCMC extracts the real spots of interest under the assumption of a Gaussian spot shape. This assumption for spot shape implies strong prior information and makes RJMCMC robust against random noise. Another benefit of RJMCMC 2DGE processing

Table 2. Spot information for the second sub-image generated from RJMCMC

num	position (row)	position (col)	width (row)	width (col)	amplitude
1	13.0000	23.0000	2.3559	3.0044	192.5783
2	13.0035	20.0101	1.9893	2.1364	182.8023
3	11.6072	22.8732	1.5061	1.5269	358.6255
4	14.0000	22.0000	3.7394	2.7883	218.3849
5	13.5044	6.5399	2.2925	1.6172	75.2568
6	5.9919	12.9131	2.3382	1.9234	22.7709
7	15.0118	5.0023	1.8144	2.1599	10.4090
8	12.2408	2.1442	2.4446	2.0642	8.5309
9	3.0066	17.4009	1.7251	1.5914	11.7475
10	3.8674	21.2892	1.6920	1.5663	9.8358
11	18.1904	11.8794	1.6128	1.5320	5.7568
12	15.0298	22.8203	1.8769	1.6150	503.9675

is that complicated spots in 2DGE from protein complexes are separated into several small spots. Moreover, the RJMCMC finds some extremely weak spots, based on Gaussian spot shape assumption, which many threshold approaches fail to detect. In addition, RJMCMC does not require that the number of spots be fixed: RJMCMC based on Monte Carlo methods searches for the proper number of spots automatically. However, the radial functions are expressed by rather large matrices, so if there are many spots in the image of interest, RJMCMC for 2DGE can be a time consuming method and possibly impracticable. Also, if the spot shape is very different from a Gaussian shape, RJMCMC in this paper will tend to generate many small Gaussian shaped spots to model the non-Gaussian spot. That is, we note that the proper spot finding may fail for non-Gaussian shaped spots. However, restoration can work even in the case of non-Gaussian spots since the overall restored shape may still be well modelled.

6 Further Work

We present the possibility of RJMCMC to process and analyze 2DGE images with a Gaussian spot shape assumption. However, it is known that the actual shape for 2DGE is non-Gaussian and non-Lorentzian. Therefore, we will incorporate more realistic shapes into the 2DGE image and this will give better Bayesian model for RJMCMC (see spots in Fig. 3). One limitation of RJMCMC for practical use is its computation time. At present, it takes 60 minutes with 10000 iterations on a Pentium CPU at 3.20GHz for the first experimental example. To make RJMCMC more practicable, sub-sampling will be applied. Next, we aim to research labeling of the spots from the RJMCMC output. Last, we will improve more sophisticated algorithms for the background subtraction as shown in Lepski's paper [4].

Acknowledgements

For ChulHun Kang and Tae-Seong Kim, this study was supported by a grant of the Korea Health 21 *R&D* Project, Ministry of Health and Welfare, Republic of Korea (02-PJ3-PG6-EV07-0002).

References

1. Reed, I. S., Yu, X.: Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, *IEEE Transactions on Acoustics, Speech, and Signal Processing*,. **38**, (1990), 1760–1770
2. Green,P.J.: Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination, *Biometrika*, (1995), 711-732.
3. Casella,G., Robert,C.P: Rao-Blackwellisation of sampling schemes, *Biometrika*, **83**, (1996), 81-94.
4. Lepski, O. V., Mammen, E., Spokoiny, V. G.: Optimal spatial adaptation to inhomogeneous smoothness: and approach based on kernel estimates with variable bandwith selectors, *The Annuals of statistics*,. **25**, (1997), 929–947.
5. Stephens,M.: Bayesian Methods for Mixtures of Normal Distributions, *PhD thesis*, Magdalen College, Oxford University, (1997)
6. Bergling, H.: Automated 2-D gel electrophoresis spot excision using Ettan Spot Picker in proteomics research , *Life Science News*. Amersham Pharmacia Biotech. (2001)
7. Godsill,S.J.: On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty, *Journal of Computational and Graphical Statistics*, **10**, (2001), 1-19.
8. Dowsey,A.W., Michael,J.D., Yang,G.: The role of bioinformatics in two-dimensional gel electrophoresis, *PROTEOMICS*, **3**, (2002), 1567 - 1596.
9. Pedersen,L.: Analysis of Two-dimensional Electrophoresis images, *PhD thesis*, Technical University of Denmark. (2002)
10. Cutler,P., Heald,G., White,I.R., Ruan,J.: A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection, *PROTEOMICS*, **3**, (2003), 392 - 401.
11. Gorg,A., Weiss,W., Dunn,M.J.: Current two-dimensional electrophoresis technology for proteomics, *PROTEOMICS*, **4**, (2004), 3665 - 3685.
12. Kaczmarek,K., Walczak,B., Jong,S., Vandeginste,B.G.M.: Preprocessing of two-dimensional gel electrophoresis images, *PROTEOMICS*, **4**, (2004), 2377 - 2389.
13. Rohr,K., Cathier,P., Worz,S.: Elastic resistration of gel electrophoresis images Based on Landmarks and Intensities, *International Symposium on Biomedical Imaging (ISBI)*, (2004), 1451-1454.
14. Bettens,E., Scheunders,P., Dyck,D.V., Moens, L., Osta,P.V.: Computer analysis of two-dimensional electrophoresis gels: A new segmentation and modeling algorithm, *Electrophoresis*, **18**, (2005), 792 - 798.
15. Yoon,J., Godsill,S.J., Kupce,E., Freeman,R.: Deterministic and statistical methods for reconstructing multidimensional NMR spectra, *Magnetic Resonance in Chemistry*, **44**, (2006), 197-209.
16. Yoon,J., Godsill,S.J.: Bayesian Inference for Multidimensional NMR image reconstruction, *European Signal Processing Conference (EUSIPCO)*, accepted, (2006).