

A Recursive Application of a Support Vector Machine for Protein Spot Detection in 2-Dimensional Gel Electrophoresis

Gary D. Boetticher [UHCL] / Hisham Al-Mubaid [UHCL] / Karen Frasier-Scott [UHCL]

Abstract

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) analysis remains the core of proteomic technology because it is currently the most powerful method for analyzing large collections of proteins. Advances in electrophoresis equipment are making this technique more accessible, but effective computer assisted protein spot detection remains a labor-intensive endeavor. Protein spot analysis is still time-consuming, requires human intervention, and is in need of further development. This paper explores a technique of recursively applying a Support Vector Machine (SVM) in identifying protein. An SVM is a powerful learner capable of optimizing differences between classes. In this context, the different classes correspond to the presence/absence of a protein. Different experiments are conducted to assess these differences in class formation in the context of a normal image and a highly saturated image.

AT PRESENT, MOST PROTEOME ANALYSIS PROJECTS BEGIN with the separation of proteins by two-dimensional electrophoresis. Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is a widely used method for separating a large number of proteins from complex protein mixtures and for revealing differential patterns of protein expressions. These protein mixtures can come from a variety of different sources such as a complex tumor tissue sample taken from a cancer patient or from a homogeneous cell tissue culture sample. An established method used to study differential protein expression is the comparison of 2-D PAGE images from different samples. A normal versus a cancer tissue type, or a stimulated vs. a non-stimulated cell tissue culture sample are examples of model systems. This type of study relies on methods that can compare images from at least two different gels. Because of the high variation between gels, detection and quantification of protein differences can be problematic.

Background

The inevitable inventory of genes that will be produced by the Human Genome Project heralds the start of a new era: the Age of Proteomics (Borman). Although DNA is the blueprint for life, it is the set of proteins that are actually transcribed and translated at any moment in time that determines the function of a particular cell. Proteomics is the study of the complete protein complement of the cell, tissue, or organism at any one time. One of the main techniques used in this growing field is two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) (Defrancesco).

Patrick O'Farrell first described 2-D PAGE technology more than 25 years ago. However, recent advances have made the technique more popular, by improving what was

often a harrowing and time-consuming experience. Fragile polyacrylamide tube gels containing drift-prone carrier ampholytes have been replaced by immobilized pH gradient (IPG) strips that allow for simultaneous "in-gel" rehydration and sample application. They offer mechanical stability due to plastic backing and provide an increased range of pH values. Most important, they improve reproducibility between gels and among different laboratories (Gorg).

The first dimension of the 2-D PAGE technique resolves proteins by isoelectric focusing (IEF). All proteins have a net charge (the sum of the charges of the amino acid side chains). The pH at which the net charge of the protein equals zero is the isoelectric point (pI). IEF separates (focuses) proteins on the basis of their charge or pI by electrophoresis across a polyacrylamide gel containing a pH gradient. Even proteins that differ from each other by only one amino acid residue can be separated in this manner. Precast, dehydrated, immobilized pH gradient gel strips are widely used for the IEF step and are available in a variety of pH ranges that can be overlapped for even higher resolution. Because the voltage necessary for IEF is generally high (up to 3,500 V) the gels are typically run horizontally on a flatbed system complete with a cooling plate.

After completion of IEF, the proteins are resolved in the second dimension via SDS-polyacrylamide gel electrophoresis (2-D PAGE). SDS not only denatures proteins, but also applies a uniform negative charge to the surface of the proteins, allowing separation based on relative molecular weight. After the IPG strips are equilibrated they are affixed to the top of a vertical SDS-polyacrylamide gel and electrophoresed. Finally, the separated proteins are visualized. Proteins labeled with isotopes are detected by autoradiography; other methods of detection commonly used are silver staining, Coomassie Brilliant blue staining, and fluorography. New stains and staining techniques now permit proteins present in low abundance to be detected in one step.

Matrix-assisted laser desorption ionization-time of flight (MALDI-ToF) mass spectrometry (MS) allows the analysis and identification of very small amounts of protein isolated from the gel. These advances have combined to make 2-D PAGE a more attractive option for the analysis of complex protein mixtures.

2-D PAGE has long been recognized as a powerful tool for the analytical analysis of biological samples. Several thousand different proteins are expressed at any one time in a particular tissue sample or organism. 2-D PAGE makes it possible to resolve these proteins. Scientists can compare samples from living samples and identify those proteins whose expression levels have changed using 2-D PAGE. With this method it is possible to distinguish functionally distinct proteins encoded by the same gene, such as mRNA splice variants and

proteins bearing post-translational modifications (e.g., methylation, glycosylation, and phosphorylation). When combined with a means to quantitatively analyze these patterns, 2-D PAGE provides the ability to analyze complex patterns of gene regulation both temporally and spatially (i.e., subcellular enrichment). The resultant two-dimensional array of spots, each of which corresponds to a single protein species, can then be analyzed with specialized computer software.

Computerized imaging and image analysis is now used to analyze the array of spots on the gel. While most standard imaging systems—storage phosphor screen or fluoro imagers, flatbed document scanners, or even old-fashioned densitometers—will do the job of collecting images of 2-D PAGE gels, specialized software has been developed for analyzing the complex patterns of spots. Sophisticated imaging and image analysis strategies are required for conclusive interpretation of two-dimensional gel electrophoresis (2-D PAGE) maps in order to identify pertinent differences in protein expression during regulation of the transcription of discrete sets of genes.

A single protein may be present as several spots on a 2-D PAGE gel due to a chemical separation of protein subunits that would normally be associated within the cell. Modified forms of a protein through post-translational processes, which change the charge of that protein, may result in even more spots representing the same protein. Many proteins do not run as a globular mass and sometimes appear as smears on the gel. There is also a major problem of physical overlapping of different proteins within the same area on the gel. Therefore, the enormous potential of 2-D PAGE is severely restricted by the difficulty of the image analysis of each of the individual proteins within the protein spots.

A few investigators, most notably Celis and his colleagues, have made remarkable in-roads in the identification of 2-D PAGE-resolved proteins and the development of 2-D PAGE maps and databases (1995), but there is considerable continued debate about the sustainability of 2-D PAGE as a platform for protein expression mapping. This debate is fueled by its technical demands and limitations that mostly stem from the analysis of the data.

Related Research

In computer-assisted Proteomic research, the comparison of protein separation profiles involves several heuristic steps, ranging from protein spot detection to matching of unknown spots. The development of more sophisticated mass spectrometry (MS)-based methods to characterize 2-D PAGE-resolved protein spots and proteins from other sources has led to an increase in the efforts now being made to exploit 2-D PAGE as a protein expression mapping tool (Celis 1998, Celis 1999, Fitcher, Hecht-Nielsen, Li, Shevchenko). Many different investigators have tried different methods to solve the technical demands and limitations of the analysis of the data.

Several current analysis programs are based on the important step of the recognition of the geometric relationship between the gel profiles, which is modeled on the basis of a given set of known corresponding spots, so-called landmarks. The locations of unknown spots are predicted using

the optimized model where efficient protein spot matching is achieved using this image-warping step. This approach is known to be incapable of modeling all the complex distortions inherent in electrophoretic data even when polynomial functions together with least squares optimizations have been used. Some investigators (Salmi) have tried to satisfy the need of more flexible gel distortion correction by using a hierarchical grid transformation method with stochastic optimization that provides an adaptive multi-resolution model between the gels to achieve automatic warping of gel images. This method seems to achieve some success in correcting the spot-matching image, but percent recognition is still too low.

The shift in research interest in recent years from Genomics to Proteomics has increased the potentials and the demand for more research work on protein activity analysis. That shift was caused by the fact that most diseases manifest themselves at the level of protein activity and the realization that genetic information seems to be insufficient to predict and distinguish healthy from diseased tissues (Kaczmarek). The focus in Proteomics is on analysis of the proteins expressed by a genome, which involves three major tasks: (1) protein spot detection, (2) protein gel image matching, (3) and spot quantitation. There are many software packages supporting these three steps. Most of them require user guidance and allow for comparison of only two gel images, but efficient Proteomics study of a disease may require the analysis and comparison of a large number of sample gels.

When investigators compared two well-known computer analysis programs (Raman) using three of the fundamental steps (spot detection, gel matching, and spot quantitation) involved in 2-D PAGE image analysis, they found each of the programs deficient in one or more of the three steps. In spot detection, even the best program only recognized 89 percent of the real spots and relatively few of the extraneous spots. When the two programs were compared using non-geometrical distortions in gel matching, both required user intervention. During spot quantitation tests, both programs did relatively well in ratios less than 1/6, but there was a large difference between the two programs using larger ratios.

One major goal of research in this domain is to minimize human intervention during the entire process to reduce the subjectivity of the human expert and to increase throughput. Other important goals include matching and alignment of several 2-D PAGE gels at a time and efficiently handling merged spots and complex regions in a 2DE gel image.

Current Methods

A number of approaches and systems have been proposed and developed for the 2-D PAGE analysis problem. Commercial software products and tools for supporting research in Proteomics are available. Among these programs are:

Delta 2D from Decodon Inc.

<<http://www.decodon.com/Solutions/Delta2D.html>>

Melanie 3.0 from Geneva Bioinformatics Inc.

<<http://us.expasy.org/melanie/Melanie.htm>>

Progenesis form Nonlinear Inc.

<<http://www.nonlinear.com>>

ProFinder 2D from Perkin-Elmer Life Sciences
<<http://las.perkinelmer.com>>
Melanie, Z3, DIGE, and CAROL.

Two sub-tasks which are universal in the gel matching process are *local matching* and *global matching*. Several of the analysis packages have attempted to build these capabilities into their software. In local matching, a number of local spots compose a pattern P in a source gel S , the goal is to compute all the local patterns P' in the target gel T that resemble both the geometric pattern and shape of P . Global matching, on the other hand, computes the overall list of spot pairs that correspond to each other in the two gels. Usually the local matching produces the landmark spots to be used for global matching.

Takashami et al. (1996, 1997, 1998, 1999) developed a fully automated set of algorithms for processing 2-D PAGE. The algorithms are based on the restriction landmark genomic scanning (RLGS) method. The implemented computer system is titled *DNAinsight*. *DNAinsight* treats RLGS profiles as pattern matching using Delaunay nets (DN) and relative neighborhood graph (RNG) algorithms. The algorithms are fully automated and the company states that no human intervention is required. One drawback of this approach is that it produces a number of false positive (ill-recognized) spots and true negative (unrecognized) spots which requires, subsequently, a time-consuming spot-revising process by a human expert. To try to overcome this drawback, Takahashi and his colleagues developed a new approach that used Gaussian modeling of landmark spots. Another algorithm called master spot pattern (MSP) has been applied to allow easily distinguishing spot patterns of diseased and non-diseased tissues.

The differential protein expression analysis, DIGE (Alban) has been used to try to improve the reproducibility and reliability of quantification between samples. DIGE includes a standard sample in each gel, thereby improving the accuracy of the protein quantification between samples.

Another approach described is one in which the spot detection algorithm depends on *watershed* transformation applied to the gradient image (Alt, Efrat, Hoffman). The analysis can interpret twin spots/streaks and so-called complex regions using a linear programming (LP) formulation. For matching, the detection algorithm performs *global-via-local* (as they call it) matching. The local matching is done in step one to generate a number of landmark spots for use in step two for the global matching. This approach is implemented in CAROL and has been deployed over the Internet for use by remote researchers. CAROL consists of two components, a core functionality component residing on a server machine, and a graphical user interface via the internet.

Kaczmarek and colleagues describe an approach for automatic matching of two 2-D PAGE images using feature-based matching technique and fuzzy alignment (FA). The FA allows automatic matching of images with different numbers of features and with unknown correspondence. The approach has been tested on real and simulated data.

Support Vector Machines

The support vector machine (SVM) (Boser, Vapnik) is one of

the most powerful classification methods and enjoys a considerable empirical and theoretical support. SVMs have proven success in many applications like object recognition (Blanz), face detection (Osuna), and text categorization (Joachims). Given data sets consisting of two types of points, positive and negative, SVM attempts to compute a separating hyperplane (a decision surface) with maximum margin between the points of the two sets. This phase constitutes the training; the computed hyperplane will be used subsequently to classify some new unseen points in the testing phase. However, the task could merely be separating the points of the given (training) data set. When various lines may be chosen as decision surfaces, the SVM method selects the middle element from the "widest" set of parallel lines. The best decision surface is determined by only a small set of (training) points called the support vectors. This case of finding the optimal separating hyperplane with maximum margin can be generalized to the non-separable cases via introducing the concept of "soft-margin" and constructing nonlinear classifiers via *kernel* functions. The kernel functions map the training data into a higher dimensional feature space, and SVM constructs an optimal separating hyperplane in the higher dimensional space corresponding to a nonlinear classifier in the input space. The kernel function can be linear, polynomial, or Gaussians (RBF). With the kernel functions and the high dimensional space, the hyperplane computation requires a quadratic programming, which is computationally intensive, and is of non-trivial implementation.

The Adatron Kernel (AK) algorithms (Anlauf, Fries) provide an alternative to this extensive computation by offering methods of finding solutions rapidly with a fast rate of convergence to an optimal solution. While the classic *Adatron* algorithm given in Anlauf was designed for linear classifiers, the KA is adapted into the (high) feature space of SVMs (Fries). The KA algorithm solves the maximization of the dual Lagrangian by implementing the gradient ascent algorithm. Finally, some of the advantages of using a support vector machine in this application include the ability to separate classes using normal at the boundary lines, and the ability to transform a curved space to linear space.

The Proposed Method

Two-dimensional gel analysis remains the core of proteomic technology because it is currently the most powerful method for analyzing large collections of proteins. Advances in electrophoresis equipment are making the technique more and more accessible to scientists previously intimidated by lengthy procedures and disappointing outcomes, but effective computer assisted image analysis of gels is still time-consuming and in need of further development. The research outlined here demonstrates an effective data-mining, machine learning process that improves the processing and automation of image analysis of 2-D PAGE gels.

The approach uses a Kernel Adatron Support Vector Machine to analyze a 2-D PAGE gel. Analyzing these images pose many challenges including:

- Proteins when stained may not be visible to the human eye either because of their low concentration, or else the stain is

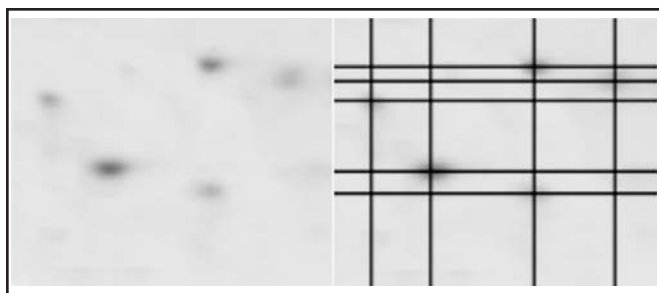


Figure 1. Original Image and Image Bounded by Centroids

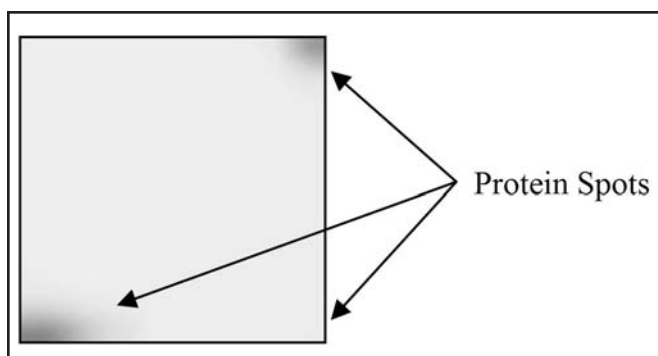


Figure 2. Sample Rectangle with three Protein Spots (two visible) (Image enlarged.)

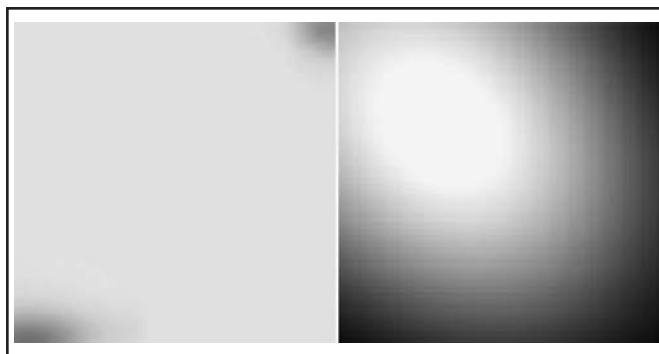


Figure 3. Original Image and Results from the Experiment 1A (Images enlarged)



Figure 4. Original Image and Results from the Experiment 1B (Images enlarged.)

not specific for that type of protein.

- Highly saturated areas make it hard to identify proteins.
- Poor image quality. A protein may smear across a 2-D PAGE gel. This can be ascribed to the chemical makeup of the protein; i.e., post-translational modification.

A traditional automated approach to analyzing a 2-D PAGE examines the whole image for detecting protein spots. However, this approach must contend with inherent image complexity in terms of pixel density diversity and the pixel topology.

An alternative approach is to preprocess the image by dynamically partitioning the image, assessing each subsection, and then merging the corresponding pieces.

Partitioning the Image

Initial analysis involves extracting pixel coordinates and the respective intensities. These intensities serve as a basis for identifying local maximas. In the event that a local maxima spans a group of neighboring pixels, then a centroid pixel is determined by using the average of the minimum/maximum values with respect to the local minima points.

These centroids act as corner anchors for extracting sub-images from the main image (see Fig. 1). Thus, any rectangle may contain up to four proteins at each of the corners. There will be no proteins at the edges; otherwise, it would be necessary to partition the rectangle into smaller rectangles. Those rectangles with zero proteins may be ignored. Figure 2 illustrates these ideas. We conducted several experiments applying the Kernel Adatron to these partitioned rectangles.

Experiment 1: Proof of Concept

All experiments seek to divide a rectangle into two classes as input values for the Support Vector Machine. Since an image contains at least two proteins, the coordinates of the local maximas for each protein will be identified as the first class. The question remains on how to identify the second class.

Since proteins occur at the corners of an image, the first experiment uses the center of the rectangle as the second class. The idea is that a single point will rapidly converge to a solution. The parameters for this experiment are set to 1000 epochs with dither set to 0.1.

Although the Support Vector Machine trained very fast, it did a poor job in recognizing the proteins as indicated in Fig. 3. Inspection of the results made it evident that more points are needed for the second class.

The next experiment further populates the second class by adding midpoints from each of the edges. The intent is to provide tighter boundaries around the points from the first class. Once again, this experiment trains for 1000 epochs with a dither rate set to 0.1.

The second experiment produced a better image than the first experiment. In this case, there was better clarity in proteins. However, as Fig. 4 illustrates, the boundaries need further definition.

The next experiment seeks to maximize the number of points in the second (non-protein) class. Maximization is accomplished by inverting the original image from Fig. 2 and extracting the maximum values. The inverted image is shown

is Fig. 5. This technique produces 14,210 points for the second class.

Figure 6 shows the results from the third experiment. The Support Vector Machine is able to identify all three proteins with crisp boundaries. A question arises regarding the extent of the boundary. The experiment strives to discriminate points into a two-class system; thus, there is a tradeoff between granularity and simplicity.

Experiment 2: Highly Saturated Region

Identifying protein spots in highly saturated regions, as illustrated in Fig. 7, is a major challenge in the field of Bioinformatics. The next experiment extracts the circled region from Fig. 7 for analysis, as depicted in Fig. 8.

Visual inspection reveals that Fig. 8 offers very little contrast. As a consequence, only four points are extracted from the second class (which corresponds to the minima values). Recall that the second class in the first set of experiments had 14,210 points.

A series of experiments are performed varying the number of epochs from 150 to 5000. The dither rate was 0.1 all the cases. The results are highly contrasted with respect to the original image. However, drawing the images closer to scale, as in Fig. 10, does not sufficiently illustrate the differences between the images. One argument is that Support Vector Machine-based image accentuates the contrast. Raising the contrast in the original image, as displayed in Fig. 11, shows that the results from the Support Vector Machine are reasonably close in appearance.

Discussion

One challenge faced by Bioinformaticists is the lack of standards in identifying protein spots. Different protein spot detection programs (also lab technicians) frequently disagree on what constitutes a protein and what is the extent, in terms of surface area, of the protein. Figure 12, a sample screenshot from an unnamed commercial product, illustrates this dilemma. The software package assumes contiguous boundaries between proteins. This assumption tends to skew the actual size of the proteins. The two-class approach of the Support Vector Machine seeks to quickly transition from where a protein ends to “non-protein” terrain.

A second challenge in defining the perimeter of a protein makes it difficult to establish a crisp border. For example, Figure 13 illustrates this situation for the two proteins in the middle of the image. One solution is to apply Bezier curves to each of the protein boundaries. Contrary to the example in Fig. 13, the Support Vector Machine produces very smooth curves with little jaggedness.

The Bioinformatic challenges, in terms of lack of standards, make it difficult to statistically assess any methodology. What seems to be important is having a defined, repeatable, and consistent process. This approach certainly offers all these features.

Conclusions

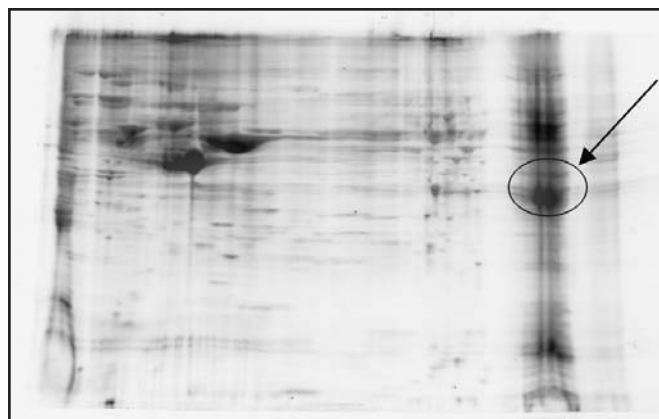
A Support Vector Machine is able to identify protein spots and mimic their shape with smooth boundaries. The approach



Figure 5. Inversion of Original Image from Figure 2 (Image enlarged.)



Figure 6. Original Image and Results from Experiment 1C (Images enlarged.)



Courtesy of the University of Texas Medical Branch (UTMB) Bioinformatics lab

Figure 7. Identified Saturated Region in a 2D Electrophoresis Gel

does very well when the original image is highly contrasted, thus producing an abundant set of points for both classes.

Furthermore, this approach is well suited for a distributed processing environment where sub-images can be delegated for processing and the corresponding results can be merged.

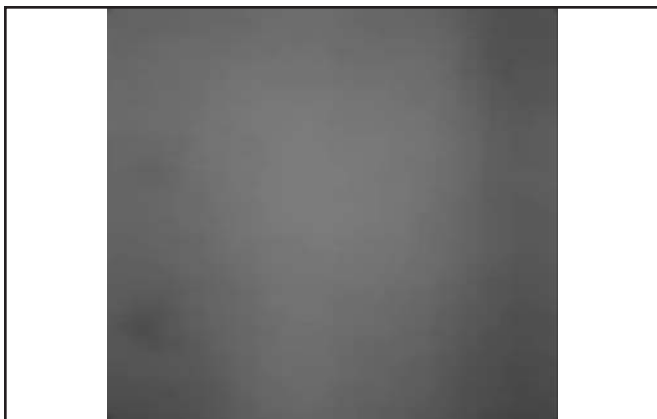


Figure 8. Enlarged Region from Figure 7

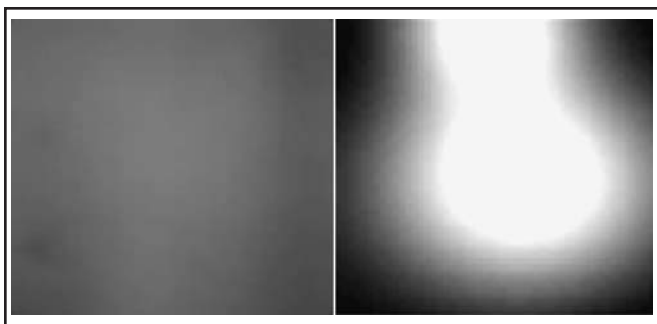


Figure 9. Original Image and Results from Experiment 2A (Images enlarged.)

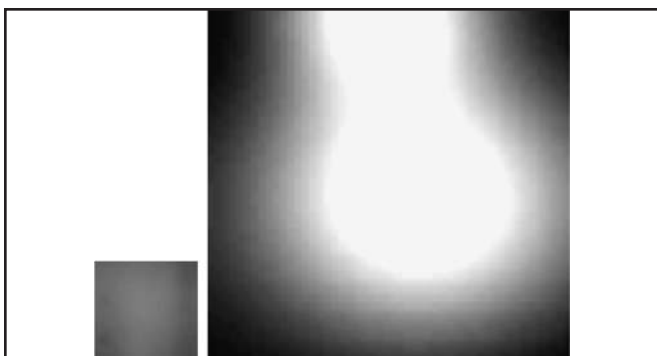


Figure 10. Original Image and Results from Experiment 2A (Sized closer to scale.)

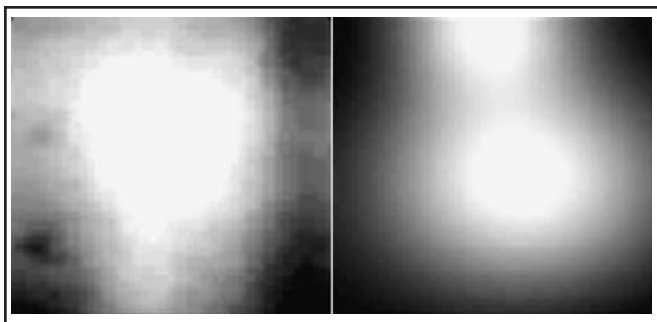


Figure 11. Original Image (Raised contrast) and Results from Experiment 2A (Images enlarged.)

Future Directions

The current approach uses a two-class system to separate spots. The benefit of this approach is simplicity in terms of representation. Adding multi-class analysis along with probabilistic reasoning would allow for refined granularity within the solution space, thus reducing the contrast within the resulting image.

It may be possible to preprocess each subregion in terms of contrast. Using a higher contrast would accentuate the differences between classes.

The approach of dynamically partitioning an image and processing each sub-image separately implies a distributed processing approach. One goal would be to develop Support Vector Machine client/server architecture. This could leverage off a distributed system and thereby provide a faster solution.

Other tasks/sub-tasks to be achieved in the future of this research include:

- Automatic spot detection with the possibility to interactively tune the results, with no, or minimum, domain expert intervention.
- Semi-automatic spot editing facilities.
- Accurate spot quantification.
- Powerful automatic spot matching and editing facilities.

Acknowledgments

Special thanks to Dr. Alex Kurosky and Dr. Allan Brasier of the University of Texas Medical Branch for the 2D Gel images.

List of Works Cited

- Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie. "A Novel Experimental Design for Comparative Two-Dimensional Gel Analysis: Two-Dimensional Difference Gel Electrophoresis Incorporating a Pooled Internal Standard," *Proteomics* 3 (2003): 36-44.
- Alt, H., F. Hoffmann, K. Kriegel, C. Wenk, and K. P. Pleissner. "CAROL—New Algorithmic Tools for Comparing 2D Electrophoresis Gel Images," Poster P21, Electrophorese Forum, Strasbourg, France, Nov. 1997.
- Anlauf, J. K. and M. Biehl. "The AdaTron—An Adaptive Perceptron Algorithm," *Europhysics Letters* 10.7 (1989): 687-92.
- Berkelman, T. and T. Stenstedt. "2-D Electrophoresis: Using Immobilized pH Gradients," *Amersham Pharmacia Biotech*, 1998.
- Blanz, V., B. Scholkopf, H. Bulthoff, C. Burges, V. Vapnik, and T. Vetter. "Comparison of View-Based Object Recognition Algorithms Using Realistic 3D Models," in *Artificial Neural Networks ICANN-96*. Berlin, Germany, 1996.
- Borman, S. "Proteomics: Taking Over Where Genomics Leaves Off," *Chemical and Engineering News* 78.31: (2000): 31-7.
- Boser, B., I. Guyon, and V. Vapnik. "A Training Algorithm for Optimal Margin Classifiers," Fifth Annual Workshop on Computational Learning Theory, 1992. 144-52.
- Celis, J. E., H. H. Rasmussen, E. Olsen, P. Madsen, H.

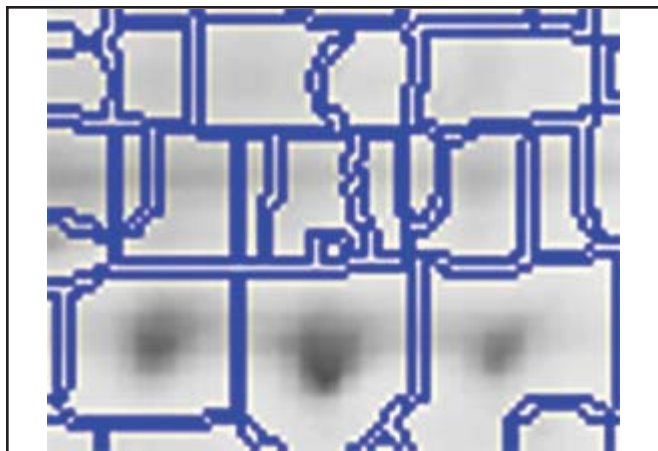


Figure 12. Image with Wide Boundaries

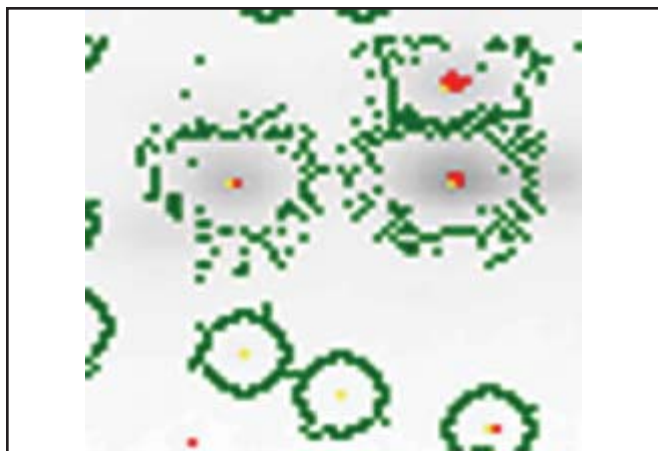


Figure 13. Image with Narrow, but Fuzzy Boundaries

- Leffers, B. Honoré, K. Dejgaard, P. Gromov, H. Vorum, A. Vassilev, Y. Baskin, X. Liu, A. Celis, B. Basse, J. B. Lauridsen, G. P. Ratz, A. H. Andersen, E. Walbum, I. Kjærgaard, I. Andersen, M. Puype, J. Van Damme, and J. Vandekerckhove. "The Human Keratinocyte Two-Dimensional Protein Database (update 1994): Towards an Integrated Approach to the Study of Cell Proliferation, Differentiation and Skin Diseases," *Electrophoresis* 15.11 (1994): 1349-458.
- Celis, J. E., H. H. Rasmussen, P. Gromov, E. Olsen, P. Madsen, H. Leffers, B. Honoré, K. Dejgaard, H. Vorum, D. B. Kristensen, M. Østergaard, A. Haunsø, N. A. Jensen, A. Celis, B. Basse, J. B. Lauridsen, G. P. Ratz, A. H. Andersen, E. Walbum, I. Kjærgaard, I. Andersen, M. Puype, J. Van Damme, and J. Vandekerckhove. "The Human Keratinocyte Two-Dimensional Gel Protein Database (update 1995): Mapping Components of Signal Transduction Pathways," *Electrophoresis* 16.12 (1995): 2177-240.
- Celis, J. E., M. Østergaard, N. A. Jensen, I. Gromova, H. H. Rasmussen, P. Gromov. "Human and Mouse Proteomic Databases: Novel Resources in the Protein Universe," *FEBS Lett.* 430.1-2 (1998): 64-72.
- Defrancesco, L. "One Step Beyond: Going Beyond

- Genomics With Proteomics And Two-Dimensional Gel Technology," *The Scientist* 13.1 (1999): 16.
- Efrat, A., F. Hoffmann, K. Kriegel, C. Schultz, and C. Wenk. "Geometric Algorithms for the Analysis of 2D-Electrophoresis Gels," *Proc.*, Fifth Annual International Conference on Computational Biology (RECOMB 2001) Montreal, Québec, Canada: ACM, 2001. 114-23.
- Evgeniou, T., M. Pontil, and T. Poggio. "Regularization Networks and Support Vector Machines," *Advances in Computational Mathematics* 13 (2000): 1-50.
- Fries, T.-T., N. Cristianini, I. C. G. Campbell. "The Kernel-Adatron: A Fast and Simple Learning Procedure for Support Vector Machines," *Proc.*, 15th International Conference on Machine Learning. Ed. J. Shavlik. Morgan San Francisco, CA: Kaufmann, 1998. 188-96.
- Futcher, B., G. I. Latter, P. Monardo, C. S. McLaughlin, J. I. Garrels. "A Sampling of the Yeast Proteome," *Mol Cell Biol.* 19.11 (1999): 7357-68.
- Gorg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. "The Current State of Two-Dimensional Electrophoresis with Immobilized pH Gradients," *Electrophoresis* 21.6 (2000): 1037-53.
- Hecht-Nielsen, R. *Neurocomputing*. Reading, MA: Addison-Wesley, 1989.
- Hoffman, F., K. Kriegel, and C. Wenk. "Matching 2D Patterns of Protein Spots," *Proc.*, SoCG'98, Minneapolis, 1998. 231-39.
- Joachims, T. "Making Large Scale SVM Learning Practical," in *Advances in Kernel Methods - Support Vector Learning*. Ed. B. Scholkopf, C. Burges, and A. Smola. MIT Press, 1998.
- Joachims, T. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proc.*, Tenth European Conference on Machine Learning (ECML-98), 1998. 137-42.
- Kaczmarek, K. "Feature Based Fuzzy Matching of 2D Gel Electrophoresis Images," *J. Chem. Inf. Comput. Sci.* 42.6 (2002): 1431-42.
- Kohonen, T. *Self-Organizing Maps*. New York: Springer-Verlag, 1997.
- Li, F. and D. C. Altieri. "Transcriptional Analysis of Human Surviving Dene Expression," *Biochem. J* 344.2 (1999): 305-11.
- O'Farrell, P. H. "High Resolution Two-Dimensional Electrophoresis of Proteins," *J. Biological Chemistry* 250 (1975): 4007-21.
- Osuna, E., R. Freund, and F. Girosi. "Training Support Vector Machines: An Application to Face Detection," *Proc.*, IEEE CVPR-97, Puerto Rico, 1997.
- Raman, B., A. Cheung, and M. R. Marten. "Quantitative Comparison and Evaluation of Two Commercially Available, Two Dimensional Electrophoresis Image Analysis Software Packages, Z3 and Melanie," *Electrophoresis* 23.14 (2002): 2194-202.
- Salmi, J., T. Aittokallio, J. Westerholm, M. Griese, A. Rosengren, T. A. Nyman, R. Lahesmaa, and O. Nevalainen. "Hierarchical Grid Transformation for Image Warping in the Analysis of Two-Dimensional

- Electrophoresis Gels,” *Proteomics* 2.11 (2002): 1504-15.
- Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann. “Linking Genome and Proteome by Mass Spectrometry: Large-Scale Identification of Yeast Proteins from Two Dimensional Gels,” *Proc., Natl Acad Sci* 93.25 (1996): 14440-45.
- Takahashi, T., K. Takahashi, M. Nakazawa, and Y. Watanabe. “High-Speed Genome Scanning Analysis Based on Automated Detection and Matching Spots in Autoradiogram Images of 2D Gel Electrophoresis,” *Genome Informatics* 7 (1996): 232-33.
- Takahashi, K., M. Nakazawa, and Y. Watanabe. “DNA Insight: An Image Processing System for 2-D Gel Electrophoresis of Genomic DNA,” *Genome Informatics* 8 (1997): 135-46.
- Takahashi, K., M. Nakazawa, Y. Watanabe, and A. Konagaya. “Fully-Automated Spot Recognition and Matching Algorithms for 2-D Gel Electrophoretogram of Genomic DNA,” *Genome Informatics* 9 (1998): 161-72.
- Takahashi, K., M. Nakazawa, Y. Watanabe, and A. Konagaya. “Automated Processing of 2-D Gel Electrophoretograms of Genomic DNA for Hunting Pathogenic DNA Molecular Changes,” *Genome Informatics* 10 (1999): 121-32.
- Vapnik, V. N. *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.