

# A New Methodology for Isolating and Diagnosing Inconsistencies in Image Matching, as Applied to the Analysis of 2-D Electrophoretic Gels

Gail Markovich, Michael Skolnick, Mark Core

Department of Computer Science  
Rensselaer Polytechnic Institute  
Troy, NY 12180

## Abstract

An image comparison algorithm employing a new notion of match consistency has been developed for the application of mutation detection on images of two-dimensional electrophoretic gels. The application requires a very high degree of accuracy in image comparison due to the rareness of mutation. The image comparison algorithm achieves high accuracy through monitoring, isolating and diagnosing inconsistencies in the matching process. The methodology is based on algorithms for monitoring symmetry relations between match hypothesis made during the course of processing.

Algorithms are given which explore violations of the basic symmetry relation. Diagnostic procedures partition symmetry violations into classes that are identified with the failure of certain essential heuristics within the comparison algorithm. This methodology provides the basis for understanding and overcoming the limitations of these heuristics in order to achieve higher accuracy.

## 1 Introduction

This paper will consider an algorithm that actively makes inferences about the locations of similarities and differences in image comparison and then attempts to bring these inferences into a consistent framework. The consistency framework allows for the algorithm's results to be monitored and analyzed to achieve higher accuracy by correcting match (or non-match) decisions. A characteristic of the algorithm is that it inherently segments problematic regions of the images being compared. This distinguishes the algorithm's approach to image matching from the widely

used methods of relaxation[1]. Another characteristic of the algorithm compared. This distinguishes the algorithms approach to image matching is that it takes a more "committed" approach to the set of potential match hypotheses, i.e., for  $n$  features being matched the algorithm entertains  $O(n)$  match hypotheses, as opposed to the  $O(n^2)$  typical for relaxation approaches.

The underlying mechanism of the algorithm is to decompose the initial correlational measures (used to assign match likelihoods) into discrete match hypotheses. (Again, the contrast with relaxation is that the correlational measure is not simply treated as an initial probability measure that will be altered via the relaxation process.) The consistency method is based on monitoring the extent to which the match hypotheses obey a basic symmetry relation. If from the "perspective" of the current match other matches look likely then, when those other matches occur, they should reciprocate their view of the current match. Thus, if "I" am a match occurring at some time step and "believe" that other matches will occur at later time steps then, those later matches should reciprocate their "belief" that "I" am a match. By tabulating these symmetry relations in a derived graph and matrix structure, the consistency methodology is able to identify regions that are in violation of symmetry, diagnose as to cause and correct the inconsistencies. This is accomplished by applying various algorithms to the matrix and graph data structures that detect the failure of thresholds and various heuristics employed in the earlier stages of the algorithm.

The methodology for achieving high accuracy image comparison will be explained on sample data of images produced from 2-D electrophoretic gels of proteins, which are being used in the detection of mutations. Mutations are of particular interest because

they are rare and thus increase the requirement for ever higher degrees of accuracy in the detection of image matches and non-matches.

The next section will describe the 2-D gel images and summarize the basic comparison algorithm. The basic comparison algorithm operates at an accuracy of about 90%. It provides the input to the consistency framework whose purpose is to isolate and repair the 10% of errors. Section 3 will present the consistency framework and results from the analysis of the framework as applied to the 2-D gel comparison problem.

## 2 The Basic Comparison Algorithm

Two-dimensional electrophoretic gels (2-D gels) are made by inducing the proteins in some object of biological interest to migrate in two-dimensions on a polyacrylamide gel. The proteins are sorted in the vertical dimension according to their molecular weight and in the horizontal dimension according to their PH. A digitized image of a 2-D gels looks like a pattern of spots (see Figure 1), where each spot represents a protein characterized by a molecular weight (vertical coordinate) and PH (horizontal coordinate). The size of a spot indicates the amount of protein. Because of the large number of proteins visualized on a typical 2-D gel (on the order of a thousand per gel), the 2-D gel technology represents a significant advance in our ability to measure bio-medical phenomena.

The operation of comparing images of 2-D gels is fundamental to the advancement of this technology. There is a special requirement for high accuracy in experimental designs where rare or small differences between gels must be detected. In the detection of mutation the problem is to detect new protein spots between experimental samples and controls. The basic algorithm for comparing 2-D gels has been applied to the detection of mutations [3, 4, 6]. This section will review the basic algorithm as applied to 2-D gels in order to provide the groundwork for a description in the next section of the new methodology [7] for isolating and diagnosing inconsistencies in image matching.

Before the 2-D gels images can be compared, both the locations and sizes of the spots on each gel image must be determined. Morphological algorithms have been used to perform this "low-level" processing of the gel images [5]. As in most low-level algorithms, a threshold operation is performed to distinguish the objects of interest from those objects judged to be insignificant. Associated with each possible spot location is a measure of the size of the spot. If the size measure is above some empirically chosen threshold



Figure 1: A digitized image of a two-dimensional electrophoretic gel of proteins.

then the spot is considered to be in the set of spots to be matched by the comparison algorithm. On the other hand, if the spot falls below-threshold then it is not considered as a spot to be matched - at least initially.

The first step in comparing the 2-D gel data is to determine the neighbors of each spot. The neighbors provide the edge relationship of the graph structure used in the matching algorithm. We have used the  $K^{\text{th}}$  Order Gabriel Graph as the basis for the neighborhood structure [2].

In the next step, the comparison algorithm determines a set of initial matches between the gel images, based on the neighborhoods. A test on a pair of spots, one from each gel, is performed by taking a cross-correlational measure (match score). As illustrated in Figure 2, potential matching spots are centered at a common origin, and the degree of clustering in the relative positions of the neighbors of the match candidates is determined by a cross-correlation measure. The relative locations of neighboring spots from the two gels are circled on Figure 2 when they "cluster" according to the cross-correlational measure. A positional noise function, applied to nearest neighbors from different gels, determines each cluster in terms of the maximal allowable positional variation in each cluster as a function of the clusters distance from the origin. Note on the bottom right of Figure 2 two neighboring spots that failed to cluster according to the positional noise function. The number of

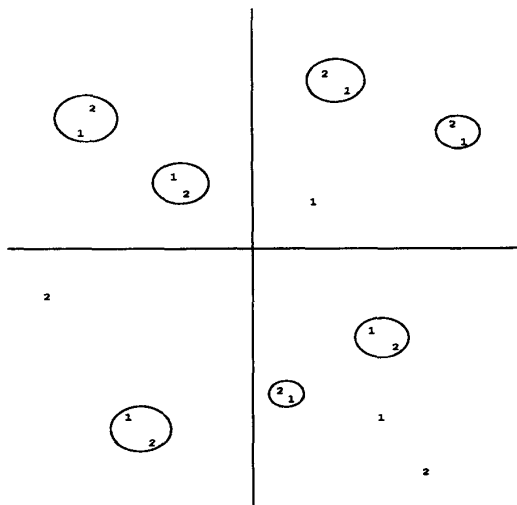


Figure 2: A cross-correlation "map" indicating the relative positions of neighboring spots from two images (1 and 2) with respect to two spots centered at a common origin. The circled neighbors are considered to "match" and contribute to the correlational measure of the pair in the center. The non-circled neighbors may be used for the possible creation of virtual spots (or nodes).

clusters determines the candidates **Match Score**. If the match score is above some (empirically chosen) threshold, then the match candidates are considered to be **Match** hypothesis, and are entered into the set of initial matches. A superposition of two non-corresponding nodes will presumably result in a random pattern and a low correlation measure. Note that the application of a positional noise function and the threshold on the match score introduces uncertainty in the match decisions; the methodology for detecting and correcting inconsistencies attempts to identify circumstances when such thresholds introduce errors.

This algorithm takes a more "committed" approach to determining and extending a set of matches. That is, an initial matching pair (seed match) is found by testing pairwise, spots from a small region on each gel. The seed match is chosen as the pair with the highest measure, and placed in the initial set. Pairs in the initial set propagate new match decisions by proposing new candidate match pairs from their neighborhoods and queueing them for testing. The neighbors that cluster by a matched pair are considered match candidates (if they have not previously been tested) or "**matches**" (read quote matches). The intuition

behind making these "matches" is that they are hypotheses in the sense that they represent local spatial proximities contributing to the match score, but if the spatial proximity is not accidental, then later on these "match" decisions should become match decisions; hence, the quotes denote a certain tentativeness.

Generally, the set of "matches" in a matching pair, will not contain all the possible "matches". Two reasons for this is that 1. spots that are above threshold on one gel may have a corresponding spot on another gel that fell below threshold, and 2. there may be genuine differences between the gel images, in which case there may be no spot to participate in a match. To understand how the comparison algorithm handles these situations, reconsider Figure 2. While the initial determination that the match candidates constituted a match depended upon the circled cluster of neighboring spots ("matches"), there remains in the cross-correlation map information about differences in the neighborhood. For example, in Figure 2 there is an isolated neighboring spot from graph 1 that formed no cluster. In order to determine whether such non-correspondences are real or not (reflect variations about a threshold or not) the image comparison algorithm searches the original image data for below-threshold spots that would reconcile the differences. Note that a heuristic has been applied which limits this below threshold search process to those isolated neighbors that are nearest neighbors of the match pairs. The heuristic embodies the fact that positional distortion increases with distance from the origin. As in the case with match score thresholds, this heuristic can sometimes cause errors, which in turn are meant to be detected by the consistency methodology discussed in the next section. There are two possible outcomes of a search for missing spots. One is that a below threshold spot is found, in which case the newly found spot can be added to the set of spots participating in the match procedure. The second possibility is that no below-threshold spot is found, in which case the comparison algorithm creates a "virtual" spot in its place. A virtual acts as a marker of differences between gels. The newly added spots can then participate in the matching process and can be considered as "match" hypotheses for further testing.

The process of queueing candidate match pairs (based on "matches") proceeds in a breadth-first search through the Gabriel graph until no new spots are visited. As the process proceeds, portions of the tree defined by the breadth-first search can continue to grow or die out. During this part of the algorithm, our

“committed” approach does not allow spots in a candidate pair that failed, to be reconsidered for matches with other spots. Also, spots in a candidate pair accepted as a match, cannot be retested with other spots to find “better” matches. Therefore, the initial set of matches contains some incorrect match pairs and “matches”, and does not contain some spots and their matches at all. This “committed” approach reduces the effect of these incorrect matches, since their branch of the tree will quickly die out as a result of propagating incorrect “matches”. Empirically we have found that this more “committed” approach results in about 90% accuracy in comparison. Additionally, a key computational advantage of this approach is only  $O(n)$  match pairs will be considered - as opposed to the  $O(n^2)$  pairs in relaxation based approaches.

The next section discusses a new graph and matrix structure created from this initial set of matches and the “match” relationship, and describes how this graph, matrix and the algorithms that operate on them, embody a new notion of consistency, which is applied to identify and correct the remaining 10% of errors.

### 3 Consistency Methodology

#### 3.1 Detection

As the basic comparison algorithm runs, a “match”/match (“m”/m) graph and matrix are created. The “m”/m graph is intended to approximate the union of the two original Gabriel Graphs, where virtual nodes have been created to account for missing nodes, and nodes in this graph have similar edge relationships as in the Gabriel Graphs. The nodes in this graph are labeled by pairs of spots (i,j) where node (i,j) is entered into the graph if node i and node j from the original Gabriel graphs were determined by the basic algorithm to match or “match”. Note that if (i,j) is involved in both a match and “match” only one node is created, with the distinction between match and “match” embodied in the graph edges. The edges are directed edges, and are defined by the “match” relationship. For example, in Figure 3, the matched pair (668,701) in the center “matched” the following pairs in its surrounding neighborhood, (654,681), (553,583), (560,590), (595,563), (766,803), (667,692), (666,698) and (663,695). Some of these pairs are matches that reciprocate this relationship, for example, the pairs (654,681), and (595,563) have edges back to (668,701). Note that only matched pairs have edges going out,

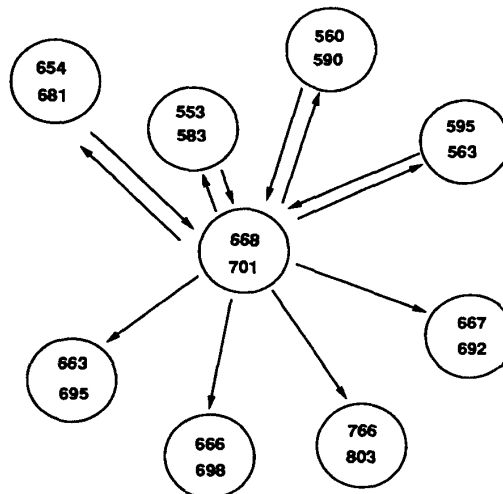


Figure 3: A portion of the “Match”/Match Graph taken from data given in Figures 4 - 6. An edge from node A to node B indicates that the spot pair in node B was “matched” (clustered) in the cross-correlational map of the spot pair in node A. Consistency is defined by reciprocating edges, A to B and B to A.

and pairs with no out edges are non-matching pairs, even though they may have “matched.”

Consistency is based on maintaining a symmetric relationship between nodes in the “m”/m graph. A pair is consistent with the current set of matches when all of the nodes adjacent to it have an edge coming back from it. For example, returning to Figure 3, (668,701) is consistent with (654,681), and (595,563) but not with (766,803), or (663,695). This symmetry requirement is derived from the following idea: earlier matches are based on “matches” (spatial correlation in the neighborhoods), which in turn should eventually become matches that reconfirm the earlier matches. If match (i,j) occurs based (in part) on “match” (i’j’) then at some latter point in the comparison process match (i’j’) should occur based on “match” (i,j). Isolating and diagnosing inconsistencies involves a systematic exploration of violations in the symmetry of “m”/m relations.

The method for exploring violations in symmetry begins by tabulating the counts of in and out edges for each node in the “m”/m graph in a “m”/m matrix (a sparse matrix which is implemented with linked lists). Each entry in the matrix contains two counts: the match score, which is the number of out edges, and the “match” score, which is the number of edges going in

to the node. The following two conditions characterize consistent symmetry between matches and "matches" in the matrix representation:

1. if the match score and "match" score of the  $ij$ th entry of the matrix are equal in value, and
2. if no other row or column coming out of  $i$  or  $j$  contains a non-zero "match" score.

Figure 4 contains a part of the matrix with both consistent and inconsistent entries. The match (660,668) is consistent according to the above criteria. The remaining matrix region represents a "coherent" set of inconsistencies in the sense that inconsistent regions are connected by non-zero row and column entries comprising small submatrices of related inconsistencies. Figures 5 and 6 provide the original gel data of the neighborhood region near the inconsistencies involving spots 766 and 803, which will be discussed in detail below. The vast majority of matches produce consistent patterns and the regions of related inconsistent rows and columns tend to be relatively infrequent and unconnected, hence limiting the scope of search through the matrix. In particular, there were 115 inconsistent submatrices involving 186 out of 1331 spots after the basic comparison algorithm was applied to the gels studied in this paper.

After locating the inconsistencies on the "m"/m matrix, the exploration of the "m"/m graph continues, with diagnosis and correction of the inconsistencies as the next step. The motivation for this diagnosis and correction algorithm can be understood by examining the patterns and classes of inconsistencies generated in the basic comparison algorithm. The diagnosis algorithm determines the pattern of an inconsistency, caused by a failure of some heuristic in the basic comparison algorithm, which then defines the corrective action to be taken.

### 3.2 Diagnosis and Correction

One class of patterns involve minor inconsistencies resulting from differences in the neighborhoods of matching spots. A basic pattern, called a **border effect**, occurs when a match is not in the neighborhood of all the neighbors of both pairs. Matching nodes may not have exactly the same pattern of neighbors at the boundary of their neighborhoods since variations on each gel cause nodes to meet the Gabriel graph edge criteria on one gel, and not on the other. In Figure 7, this can be found in the upper right quadrant: node 1; matches node 2; but 2; is outside the neighborhood of  $2_a$  so in the match  $(1_a, 2_a)$ , 1; is "matched" with 2;.

	668	692	695	697	698	701	803	1364
660	20							
663	"20"							
664		"1"						
665			15					
666			"14"					
667				14				
668				"13"	"1"			
669					15			
766		"2"			"14"			
803		"11"				"1"		
1364						19		
660						"14"		
663							14	
664						"2"	"1"	
665								12
666							"12"	"1"

Figure 4: An example of a portion of the "match"/match matrix. Numbers on the left column represent labels of spots from one image (Figure 5), and numbers on the top row represent labels of spots from another image (Figure 6). Entries may contain match scores, and "match" scores. The submatrix contains a consistent entry, (660,668) and the rest of this matrix is an inconsistent submatrix.

A related pattern called a **shift effect** occurs when the location and local disparity of neighboring nodes cause them to be incorrectly "matched" by the nearest neighbor heuristic. This is a fundamental limitation of the noise function to capture higher order patterns of noise. In Figure 7, neighbor  $2_x$  and  $1_y$  are nearest neighbors and are incorrectly "matched", where  $(1_x, 2_x)$  and  $(1_y, 2_y)$  are the correct matches. Both of these patterns can be found in the "m"/m matrix as a row or column with more than one "match" entry where this entry has a small value.

More severe inconsistency patterns can occur as a result of the "committed" nature of the basic algorithm. Any of the minor inconsistencies described above may propagate more errors. Because the basic comparison algorithm queues a node only once, if the first time a node is "matched" is incorrect, e.g. due to a shift or border effect, a more serious inconsistency results. The incorrect "match" is queued and the match score is determined, with the result that the "match" can be accepted as a match, or fail to match. If the "match" is accepted it is a **mismatch**, and if it fails, both nodes in the "match" will not be able to participate again in the queueing (the nodes are **blocked**). Both of these patterns also appear in

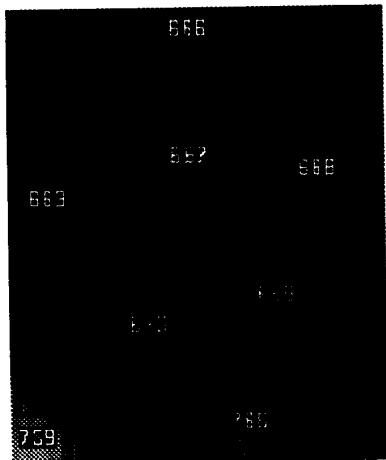


Figure 5: Gel image containing part of the inconsistent region near spot 776 from one gel.

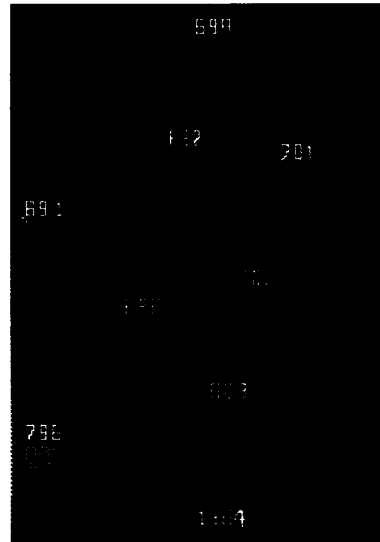


Figure 6: Gel image containing part of the inconsistent region near spot 803 from other gel.

a submatrix, with a row or column containing more than one “match” entry, but with the entry of the correct match receiving the most support (“match” score) from its surrounding matches.

The diagnosis and correction algorithm is designed to work in three passes executed sequentially (see Figure 8), with each pass correcting a different type of error. The more severe errors involving mismatches and blocked matches are corrected first. The second pass assumes all matches have been created correctly between pairs of above threshold spots, and determines when to search for below threshold spots or create virtuals. In the third (and last) pass, any minor inconsistencies are corrected, e.g., the simple border effects or shift patterns. Because many patterns may exist in a single submatrix, each pass of the algorithm is applied to a single row or column at a time. To ensure that this very local approach to diagnosis and correction maintains or increases the overall consistency of the results, the requirement of symmetry between pairs of matching nodes is embedded in each pass of the algorithm, i.e., any change contemplated must not decrease the symmetric consistency.

As the algorithm proceeds, the “m”/m graph and matrix are continuously updated with any newly created matches or nodes. The end results should be a totally consistent matrix with exactly one entry in each row or column. Note that a totally consistent matrix is in fact achieved when the diagnosis and correction algorithm is applied to the gels studied in this

paper. The remainder of this section examines the diagnosis and correction algorithm as it is applied to a section of actual gel data. We will consider the steps that altered the initial submatrix from Figure 4.

The inconsistent submatrix of Figure 4 contains an example of many of the patterns of inconsistencies described above. Notice the entries (766,1364) and (669,803) are matches, with low “match” scores (meaning very few other matches viewed them as “matches”), and there is a “match” entry for (766,803) with a high score. The correct match should have been (766,803). The error in the basic algorithm occurs when the “match”(669,803) is queued, and then becomes a match as a result of enough correlation in its neighborhood. In turn, 766 becomes an isolated spot with respect to the match of (669,803), which results in the incorrect creation of virtual node 1364. In addition, row 667 is one example of a blocked match pattern, and the other inconsistent “match” entries are shift and border effects.

Pass 1 of the consistency algorithm begins in this region by modifying row 667. The best match for the row is determined to be between 667 and 692 by comparing the cross-correlational measure for pairs 667,692 and 667,701. There is no match in the column of 692, and the match 667,692 is recorded. Later passes will correct the “match” of 667 and 701. The

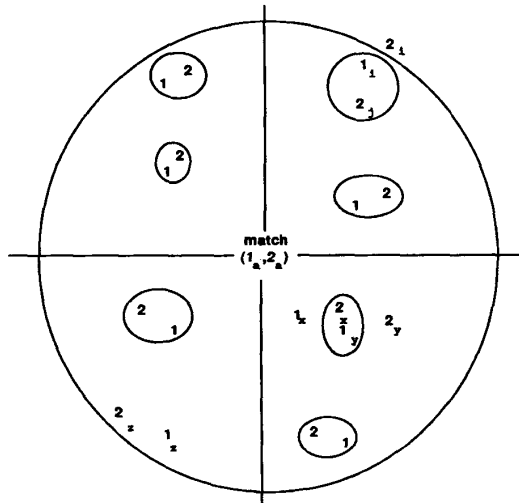


Figure 7: Neighborhood illustrating simple inconsistencies. Indicating basic patterns of noise creating 1. variations at the boundary of neighborhoods, border effect (subscripted by  $i$  and  $j$ ), 2. shift effects (subscripted by  $x$  and  $y$ ) and 3. complete failure to cluster (subscripted by  $z$ ).

row 669 is unchanged when analyzed by the first pass, since 669,803 is determined to be the best match for this row, out of the "match" entries in this row. The inconsistency in this row is corrected in the second pass, when spot 702 is raised above-threshold to match with 669. The row 766 is modified next. Here, the cross-correlational measure determines that 766 and 803 is an acceptable match along with the current match for this row. Since this pair also conflicts with the match 669,803 in the same column, its consistency must also be compared with 669,803. By examining the amount of symmetry in all matches, 766,803 is determined to be the best match for the column 803 and the row 766, and this match is recorded. The two conflicting matches, 766,1364 and 668,803 are removed. Virtual node 1364 is no longer needed and it is also removed from the data structures. Figure 9 shows the matrix after the above changes from pass 1.

Pass 2 of the algorithm assumes all matches have been found between pairs of above threshold spots and have been determined correctly. This pass then searches for below threshold spots, and creates virtuals when necessary. In this region, below threshold spots are found and matched with rows 663 and 669. The remaining "match" entries are shifts or border effects that are corrected in Pass 3 of the algorithm and this

- PASS 1** (mismatches and blocked matches)  
 for each row with more than one "match" entry  
 for all "match" entries in the row  
 determine above threshold matches  
 determine best match  
 i.e. pair with the highest symmetry measure  
 remove any conflicting matches in same row or column
- PASS 2** (below-threshold and virtuals)  
 for each row or column with no match  
 determine nearest neighbor  
 search based on nearest neighbor  
 if a node is found  
 try the match  
 if a node is not found  
 create virtual to match
- PASS 3** (shifts, border effects, and simple thresholds)  
 for each row with more than one "match" entry  
 determine from the "m"/m graph the match contributing to the "match"  
 check if node of this row and its match are in the neighborhood  
 add if missing from the neighborhood and reconstruct both matches  
 check nearest neighbor heuristic falling  
 override nearest neighbor heuristic

Figure 8: Diagnosis and Correction Algorithm. Note that all actions taken are required to not degrade the "match"/match symmetry relation. The classes of errors addressed by each pass are indicated in parentheses.

region in the matrix then becomes completely consistent.

This example demonstrates the data structures and algorithms that are used in applying the consistency methodology to isolating, diagnosing and correcting inconsistent match decisions of the basic algorithm described in the previous section. In summary, diagnosis is achieved by using the matrix to identify related inconsistencies and exploring the graph structures to determine the causes of the inconsistency. Consistency is improved by measuring the consistent symmetry of matches to determine when corrections are necessary.

## 4 Conclusion

A new "m"/m method for detecting inconsistencies in image matching has been described. It has been applied to the problem of detecting differences in protein spots on 2-D Electrophoretic gels, with regard to the biomedical application of detecting rarely occurring image differences due to mutations. The appli-

	668	692	695	697	698	701	803
660	20 "20"						
663		"1"					
664			15 "15"				
665				14 "14"			
666					15 "15"		
667		12 "12"				"1"	
668						19 "19"	
669						"2"	
766							17 "17"

Figure 9: The same submatrix as in Figure 4 after Pass 1 of the diagnosis and correction algorithm is applied.

cation requires a very high degree of accuracy and, we believe, calls for a new algorithmic approach to understanding the nature of high accuracy. Beyond this particular application, the methodology should be of interest in the general image comparison problem since it represents an alternative approach to the widely used relaxation methods.

The "m"/m method starts by taking a more "committed" approach to the matching process. Moving out from some initial match between a pair of nodes the spatially clustered neighboring pairs of nodes are used both to determine the correlation measure (or match score) and are placed on a queue of next match candidates. In this fashion the basic comparison algorithm only considers  $O(n)$  match pairs in a breadth-first search manner. On the images of the current application, this simple and "committed" procedure typically results in about 90% accuracy in comparisons. The problem then remains of how - in a computationally reasonable and complete fashion - can the 10% of errors be filtered out of the majority of correct decisions.

In order to avoid extensive backtracking over the majority of correct decisions, a method to detect inconsistencies in match hypotheses has been developed. The measure of consistency is based on detecting violations in the symmetry of match hypotheses. Tabulation of the match hypothesis into two data structures have been used to isolate, diagnose and correct inconsistencies due to symmetry violations: a "m"/m pair graph, and a "m"/m matrix. Diagnostic procedures that operate on classes of inconsistencies within the

"match"/match data structures can be identified with certain essential heuristics within the comparison algorithm. Each diagnostic procedure can be considered an "anti-heuristic" in that it tracks down failures of its corresponding heuristic.

## 5 Acknowledgments

This research was supported in part by NIH grant NS21219 and USPHS grant S07RR07104. The 2-D gel application and data is part of a collaboration with Drs. Samir Hanash and James V. Neel of the University of Michigan. Thanks go to the RPI Undergraduate Research Program for support of student time.

## References

- [1] Davis, L. S. and Rosenfeld, A., "Cooperating processes for low-level vision: a survey", *Artif. Intell.*, Vol 17, 1981, pp. 245-263.
- [2] Gabriel, K. R. and Sokal, R. R., "A new statistical approach to geographic variation analysis", *Syst. Zool.*, Vol 18, 1969, pp. 259-270.
- [3] Kuick, R. D., Skolnick, M. M., Hanash, S. M. and Neel, J. V., "A two-dimensional electrophoresis-related laboratory information processing system: Spot matching", *Electrophoresis* Vol 12, 1991, pp. 736-746.
- [4] Skolnick, M. M. and Neel, J. V., "An algorithm for comparing 2-D electrophoretic gels, with particular reference to the study of mutation", *Advances in Human Genetics*, Vol 16, H. Harris and K. Hirschhorn, eds., Plenum Press, New York, 1986, pp. 55-160.
- [5] Skolnick, M. M., "Morphological algorithms for the automatic analysis of 2-D electrophoretograms", *Computer Vision, Graphics and Image Processing*, Vol 35, No 3, 1986, pp. 306-332.
- [6] Skolnick, M. M., "Automated comparison of 2-D gels", *Proc. of the IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Vol 35, No 3, 1986, pp. 306-332.
- [7] Skolnick, M. M., "A general methodology for achieving high accuracy in image comparison, with application to the analysis of two-dimensional electrophoretic gels", *SPIE Conf. Applications of Artificial Intelligence VIII*, Vol 1293, 1990, pp. 536-547.