

2-D ELECTROPHORESIS GEL REGISTRATION USING FEATURE MATCHING

Mike Rogers, Jim Graham

Imaging Science and Biomedical Engineering,
University of Manchester, UK
mike.rogers@man.ac.uk

Robert P. Tonge

Protein Science,
Enabling Science and Technology
(Biology), AstraZeneca, UK.

ABSTRACT

We have carried out an evaluation of feature matching algorithms and developed an automatic method of 2-D Electrophoresis gel registration. We show that bi-partite graph matching is not appropriate in this context, and have used a combination of shape context and euclidian distance measures together with closest point correspondence estimation in our scheme. Our final algorithm is shown to be more robust than a recently published image intensity based method.

1. INTRODUCTION

Recently, proteomics research has become a large growth area in the bio-sciences. Often, studies involve differential analysis of large sets of 2-D Electrophoresis (2-DE) gels. Figure 1 shows examples of sections from such gels. In order to carry out such an investigation, correspondence between spots on sets of gel images must be determined. This implies that a transformation relating one gel image to another is required. The production of 2-DE gels is inherently variable. As a result complex non-linear deformation are often required to align comparable gels. These deformations are often difficult to identify manually and time-consuming to correct. The goal of this work is to develop a feature matching scheme that can be used to bring pairs of gels into alignment.

2. RELATED WORK

A commonly used feature matching scheme is the Iterative Closest Point (ICP) algorithm [1]. ICP iterates between estimating correspondences and then determining transformation parameters given this correspondence. Standard ICP uses a euclidian distance metric with a closest point method of determining correspondence. Within the same algorithmic framework, Belongie *et al* [2] proposed an alternative *shape context* distance metric, with optimal Bi-Partite Graph Matching (BPM) determining correspondence. Alternatives involving more complex distance, correspondence and search schemes [3][4] have been suggested but will not be evaluated here.

Recently, there has been interest in using image registration techniques for automatic gel alignment. Smilansky [5] used a block matching based approach with transformations varying from rigid to thin plate splines. The algorithm, although parameterless, contains many heuristic assumptions that may not hold for every gel alignment. Veaser *et al* [6] used image derivative information to move the control points of a bilinear interpolating grid transform. The algorithm was set in a multi-resolution framework, with a progressively finer grid of control points. The choice of a bilinear interpolating transformation does not produce a smooth deformation and can lead to unrealistic transformations. Also, the choice

of a grid of control points may not be able to represent adequately the underlying deformation, information about which is contained in spot positions.

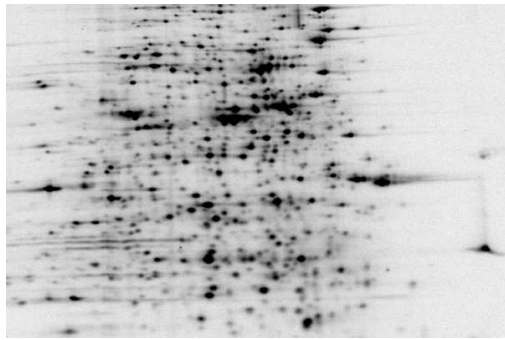
3. FEATURE MATCHING

We wish to determine the underlying *deformation* relating two gels using detected spot centres as features. There are often important, genuine, differences in the patterns of spots in gels being aligned. As well as these differences, automatic spot detection algorithms often introduce additional spurious spots whilst missing or, more commonly, merging overlapping spots. These factors will create *outlier* points in each gel of a comparable pair that have no counterpart in the other. These outliers typically lie within the range of the true spot pattern.

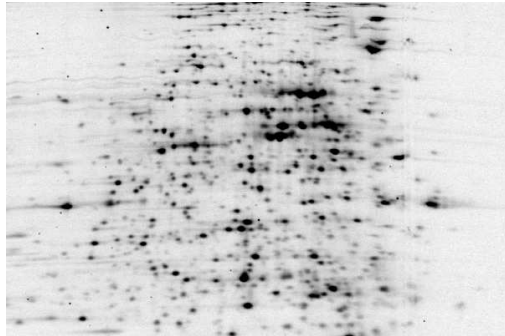
With this in mind, we have evaluated several candidate feature matching algorithms in terms of their sensitivity to both *shape deformation* and the presence of unmatchable *outliers*. We have based our investigation on the Iterative Closest Point (ICP) algorithm [1], modifying the basic scheme with alternative distance and correspondence methods. We have compared euclidian (Euc) and Shape Context (SC) [2] distance measures, together with Closest Point (CP) and Bi-partite Graph Matching (BGM) to determine correspondence. We have used a fast implementation of the unit-cost transportation algorithm [7] to solve the BGM. The permutations of these methods gives us a total of 4 algorithms, denoted: Euc-CP, Euc-BGM, SC-CP, SC-BGM. Previous work has only considered either Euc-CP [1], or SC-BPM [2] variants. As gel images are often related by a non-rigid transformation, we have used a thin plate smoothing spline instead of an affine transformation to parameterise deformation. In general, thin plate splines do not yield diffeomorphic transforms. However, we have chosen spline smoothing parameters that make transformation folding or singularities sufficiently for this data. Diffeomorphic transformations could be guaranteed using more computationally expensive spline formulations [8]. Points estimated to be in correspondence at each iteration were used as the control points of the spline. The smoothing parameter was set at 0.5, allowing non-rigid, but globally smooth deformations to be well represented.

The algorithms were evaluated using synthetic point sets, referred to as the source and target sets. Each source set consisted of 100 points drawn randomly from a uniform distribution within the unit square. Each target set is copy of the corresponding source set, giving us a ground truth correspondence. Outliers can be introduced to either set by removing or adding points in the target at random. Varying degrees of shape deformation can be introduced by transforming the target set using a thin plate spline with random control point offsets of restricted magnitude.

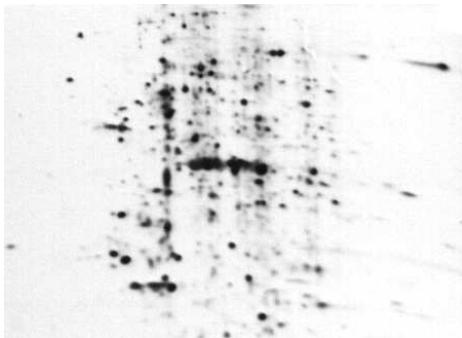
The sensitivity of the algorithms to outliers *either* in the source



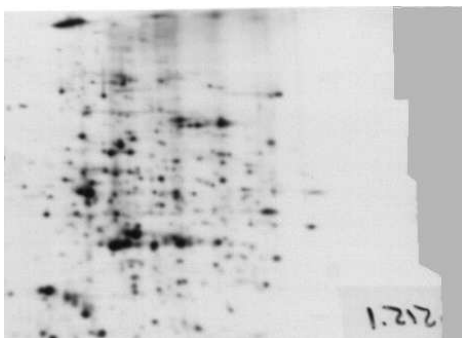
(a) 1a



(a) 2a



(c) 8a



(d) 8b

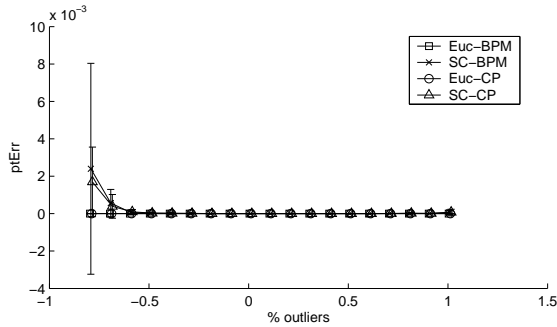
Fig. 1. Examples sections from 2-DE gel images from the Z3 set (see section 4).

or target set was evaluated. Target and source sets were created with the proportion of outliers varying between -0.8 (80% of points removed) and 1 (100% extra points added). Additionally, the amount of shape deformation added was varied within the range observed between comparable gels; from almost perfectly aligned (no deformation) to large, complex deformations (offsets of up to 25% for each random control point). Each algorithm was used to align 50 random source and target sets at each outlier and shape deformation setting. Figure 2 shows the mean-squared euclidian distance (± 1 s.d. error bars) between ground-truth corresponding points after alignment. A distance of 0 indicates a perfect alignment of corresponding points in both sets. With no shape deformation (figure 2(a)), both Euc algorithms succeed in aligning the (co-incident) sets regardless of outlier proportion. In contrast, both SC algorithms fail when large number of points are removed ($< -50\%$) even when no deformation has been applied. However, SC is most insensitive to the amount of shape deformation (figures 2(b)-(c)) applied, whilst both Euc methods begin to fail as deformation increases. With large deformation (figure 2(c)), BGM is shown to reduce the error introduced by shape deformation, but only when the numbers of outliers are relatively small (approx. -20% to 10%). This experiment suggests that either SC algorithm is a good choice for gel alignment unless an extremely large number of outliers exist in the source set.

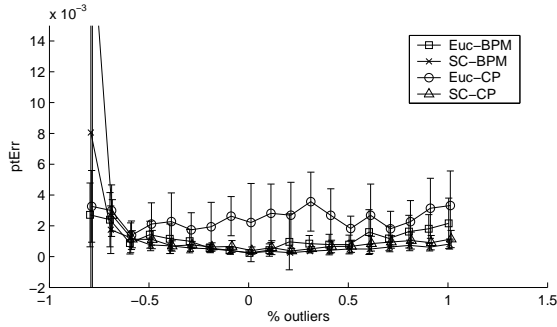
The previous experiment is not entirely realistic in the context of gel alignment, as outliers are likely to occur in both sets simultaneously. We repeated the above evaluation, this time *both* removing original points *and* adding extra points to the target set. Point sets were created with outlier proportions varied between 0 (0 points added and removed) and 0.8 (80% points removed, 80% added). Figure 2 shows the mean-squared euclidian distance (± 1 s.d. error bars) between ground-truth corresponding points after alignment. Surprisingly, even with no deformation (figure 3(a)) both BGM algorithms degrade as the number of extra and missing-point outliers increases. Inspection of these results shows "chains" of incorrectly assigned, but relatively consistent correspondences which introduce error into the global transformation. This affect is only observed when the target set contains both missing *and* extra points, and has not been reported elsewhere. The SC-CP algorithm is also sensitive to outlier proportion (increasingly $> 30\%$). Euc-CP is unaffected by outliers when no shape deformation is present. However, as deformation increases (figures 3(b)-(c)) the performance of the Euc-CP algorithm quickly degrades whilst the performance of the other three degrade more slowly.

Although we are assuming the presence of outliers, none of our algorithms explicitly address this¹. We introduce robustness to outliers by using M-estimators [9] in the calculation of the transform for a given correspondence. M-estimators iteratively down-weight point pairs which are inconsistent in euclidian distance with the overall distribution of point pair distances. In this way, outlier correspondences receive less weight in the calculation of transformations. The effect of using M-estimators in this scheme is shown in figure 4. Comparing figure 4 with its non M-estimator counterpart (3(b)) shows that M-estimators reduce error for each of the algorithms (Euc-CP to a lesser extent). Both BGM algorithms are improved to approximate SC-BGM up to 35% outliers, but between 35%-50% -BGM still gives the best performance. M-estimators are a valuable addition to our feature matching scheme.

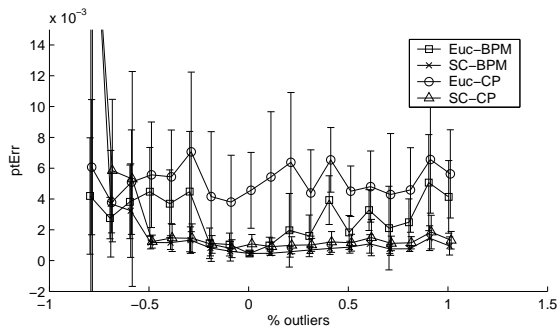
¹SC includes a heuristic estimation of outliers



(a) No deformation



(b) Medium deformation

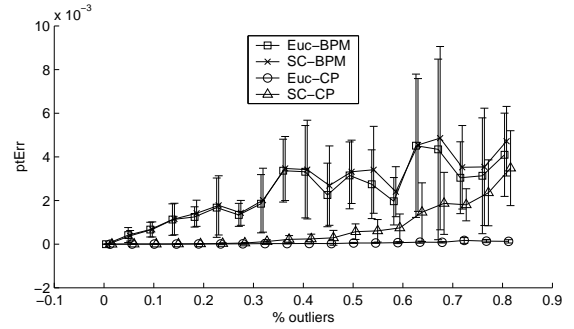


(c) Large deformation

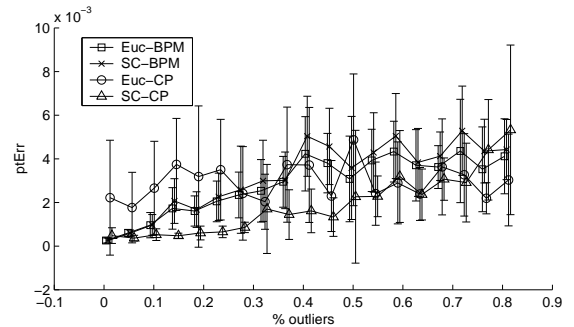
Fig. 2. Effects of additional *or* removed outliers on feature matching algorithms with varying amounts of deformation.

4. GEL REGISTRATION

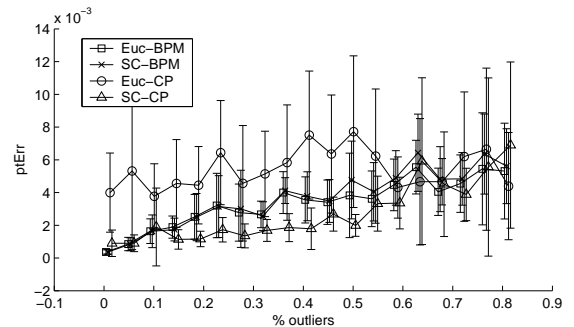
We have used the feature matching results presented in section 3 to design a gel registration scheme. As there is information in the images useful for alignment at various scales, we have use a multi-resolution approach. At each level of image resolution, spots centres are detected from both gels using a simple threshold on the Laplacian. The resulting point sets are used to determine a thin plate spline transform using a combination of Euc-CP and SC-CP algorithms with M-estimators. Figure 3(a) shows that Euc-CP is most appropriate when shape deformation is small, and figure 3(c) shows that SC-CP performs best when deformation is larger. For this reason, we have adopted a scheme that uses a weighted average of Euc and SC as a distance measure. The weights chosen vary linearly between entirely SC when deformation is expected to be large (lowest resolution), to entirely Euc when remaining deformation is expected to be small (highest resolution). In common



(a) No deformation

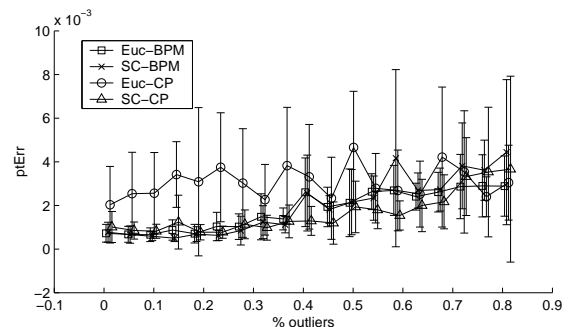


(b) Medium deformation



(c) Large deformation

Fig. 3. Effects of additional *and* removed outliers on feature matching algorithms with varying amounts of deformation.



Medium deformation, using M-estimators

Fig. 4. Effects of M-estimators on feature matching algorithms with non-rigid deformation and varying additional *and* removed outliers.

with other gel registration algorithms [5][6] we have also adopted a coarse-to-fine smoothing scheme, varying the smoothing parameter of the thin plate spline linearly between 1 (affine, lowest resolution) and 0.05 (allowing fine distortions, highest resolution).

We have applied this algorithm to the set of images used by Smilansky [5]. The set consists of 13 image pairs² and contains a wide variety of gel staining methods, substance types and image qualities, some with many artifacts and noise. The set also varies in the difficulty of alignment, from manually difficult to align (8a and 8b, figure 1(c) and (d)) or large single dimensional deformations (10a and 10b) to being in almost perfect initial registration (11a and 11b).

We have evaluated our registration results in terms of the following mean squared intensity difference residual: $r = \sum (I_p^{t'} - I_p^{r'})^2 / n_p$ where p is the set of all pixels in the centre 50% of the image, n_p is the number of pixels in this set, $I^{t'} = I^t - \sum (I_p^t) / n_p$ is the pixel intensity of the target image I^t normalised by average pixel intensity over p , and $I_p^{r'}$ is the registered image, normalised in the same way. A residual of 0 indicates a registration result where every corresponding pixel has the same intensity.

We have compared our algorithm to that of the intensity based-block matching algorithm of Smilansky [5]³ using 11 of the image pairs. Two registrations were carried out on each pair, using one image as target and one as source and vice-versa. We present residuals from both methods together with an indication of the success of the registration. A registration was judged to be successful if it was estimated that less than 25% of the spots in the whole of the images were misaligned. This was determined manually by inspection. Results are shown in table 1. Images for each of the registrations are available on the web⁴.

Smilansky's method fails six times in total. Each of these failures is large and clear-cut. Feature matching fails on only 2 pairs (8a and 8b both ways). 8a and 8b are particularly hard to align manually, so failure here is not surprising. Where registrations have been successful, the residuals obtained by each method are similar. The exception to this is the 7a-7b pair which results in a larger residual using feature matching. This alignment is a borderline failure with some spots in the body of the image misaligned. These results show that the feature matching gel registration algorithm produces image residuals that are equivalent to those produced by an image intensity based algorithm despite using only point features. The feature matching algorithm is also more robust.

5. CONCLUSIONS

In this paper we have presented a comparative evaluation of point matching approaches in the context of gel image registration. Evaluating all combinations of Euc,SC,CP and BGM has shown that BGM based approaches fail when both additional and missing outliers occur in the data. This finding had not been reported in the original reference [2]. Bearing this in mind, a gel registration algorithm was designed using a combination of shape context and euclidian distance with a closest point correspondence estimator. M-estimators were used to further increase the robustness of the system to outlier features. The registration algorithm was shown to achieve equivalent accuracy and to be more robust than an image intensity based registration algorithm, despite only using point features.

²3a, 3b, 4a and 4b not included due to image loading difficulties.

³Results obtained using Z3 v3.0.7 evaluation version

⁴<http://www.isbe.man.ac.uk/~mdr/gelreg.html>

Gel Pair	Smilansky		Feature Matching	
	Success	r	Success	r
1a-2a	✓	0.0122	✓	0.0120
2a-2b	✓	0.0049	✓	0.0049
5a-5b	✓	0.0031	✓	0.0031
6a-6b	✓	0.0041	✓	0.0040
7a-7b	✓	0.0049	✓	0.0073
8a-8b	×	0.0326	×	0.0447
9a-9b	✓	0.0018	✓	0.0022
10a-10b	×	0.0138	✓	0.0047
11a-11b	✓	0.0001	✓	3×10^{-5}
12a-12b	×	0.0385	✓	0.0002
13a-13b	✓	0.0117	✓	0.0112
2a-1a	✓	0.0134	✓	0.0132
2b-2a	✓	0.0043	✓	0.0049
5b-5a	×	0.0166	✓	0.0032
6b-6a	×	0.0233	✓	0.0097
7b-7a	✓	0.0064	✓	0.0069
8b-8a	×	0.0252	×	0.0468
9b-9a	✓	0.0025	✓	0.0025
10b-10a	✓	0.0041	✓	0.0045
11b-11a	✓	9×10^{-4}	✓	3×10^{-5}
12b-12a	✓	0.0043	✓	0.0002
13b-13a	✓	0.0103	✓	0.0117

Table 1. Post-registration residual for Z3 and Feature Matching.

6. REFERENCES

- [1] P J Besl and H D McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Vision*, vol. 14, no. 2, pp. 239–256, 1992.
- [2] S Belongie, J Malik, and J Puzicha, "Shape matching and object recognition using shape context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, April 2002.
- [3] A Ranagarajan, H Chui, E Mjolsness, S Pappu, L Davachi, P Goldman-Rakic, and J Duncan, "A robust pattern matching algorithm for autoradiograph alignment," *Medical Image Analysis*, vol. 1, no. 4, pp. 379–398, 1997.
- [4] A Rangarajan, H Chui, and J S Duncan, "Rigid point feature registration using mutual information," *Medical Image Analysis*, vol. 4, pp. 1–17, 1999.
- [5] Z Smilansky, "Automatic registration for images of two-dimensional protein gels," *Electrophoresis*, vol. 22, pp. 1616–1626, 2001.
- [6] S Veese, M J Dunn, and G-Z Yang, "Multiresolution image registration for two-dimensional gel electrophoresis," *Proteomics*, vol. 1, pp. 856–870, 2001.
- [7] H Achatz, P Kleinschmidt, and K Paparrizos, "A dual forest algorithm for the assignment problem," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 4, pp. 1–11, 1991.
- [8] C Twining, S Marsland, and C J Taylor, "Measuring geodesic distances on the space of bounded diffeomorphisms," in *BMVC*, 2002, vol. 2, pp. 847–856.
- [9] P J Huber, *Robust Statistics*, Wiley, New York, 1981.