

Two-dimensional generalisations of dynamic programming for image analysis

C.A. Glasbey
Biomathematics and Statistics Scotland
King's Buildings, Edinburgh, EH9 3JZ, Scotland

April 3, 2008

Abstract

Dynamic programming (DP) is a fast, elegant method for solving many one-dimensional optimisation problems but, unfortunately, most problems in image analysis, such as restoration and warping, are two-dimensional. We consider three generalisations of DP. The first is iterated dynamic programming (IDP), where DP is used to recursively solve each of a sequence of one-dimensional problems in turn, to find a local optimum. A second algorithm is an empirical, stochastic optimiser, which is implemented by adding progressively less noise to IDP. The final approach replaces DP by a more computationally intensive Forward-Backwards Gibbs Sampler, and uses a simulated annealing cooling schedule. Results are compared with existing pixel-by-pixel methods of iterated conditional modes (ICM) and simulated annealing in two applications: to restore a synthetic aperture radar (SAR) image, and to warp a pulsed-field electrophoresis gel into alignment with a reference image. We find that IDP and its stochastic variant outperform the remaining algorithms.

Key words: Forward-Backwards Gibbs Sampler, image restoration, image warping, Markov random field, pulsed-field gel electrophoresis, simulated annealing, synthetic aperture radar.

1 Introduction

Dynamic programming (DP) is a fast, elegant method for finding the global solution to a class of one-dimensional (1-D) optimisation problems. For example, it can be used to find maximum *a posteriori* (MAP) estimators of 1-D boundaries, to automatically segment 2-D images (Glasbey and Young, 2002): to find a boundary from the top to the bottom of an image, we seek the vector of integer values (f_1, f_2, \dots, f_n) that minimises

$$C(f) = \sum_{i=1}^n [Y_{i,f_i}^* + \lambda(f_i - f_{i-1})^2]. \quad (1)$$

Here Y_{ij}^* denotes an element in an $n \times m$ array, the cost of the boundary passing through location (i, j) , indexed by row i , column j and λ is a non-negative constant. (Here, and subsequently, to simplify the notation we do not explicitly exclude terms involving f_0 although it is to be understood that they are defined to be zero in all cases.) The cumulated cost, C , can be regarded as either a penalised log-likelihood or a Bayesian log-posterior density, with the smoothness of the boundary determined by the magnitude of λ . The Principle of Optimality (Bellman, 1957) is satisfied by C , so the global minimum can be found elegantly and efficiently by DP. Hidden Markov Models (MacDonald and Zucchini, 1997) yield similar log-likelihoods to (1), and are maximised using the Viterbi algorithm, a form of DP.

Further, DP can be used for 1-D image warping, also termed dynamic time warping. In order to genotype bacterial samples, gel columns, such as those in the pulsed-field gel electrophoresis (PFGE) shown in Fig 1(a), need to be warped into alignment with columns in a reference database, such as Fig 1(b) (Glasbey et al., 2005). DP finds the warping, the vector of values (f_1, f_2, \dots, f_n) drawn from a finite set, denoted \mathcal{F} , that minimises

$$C(f) = \sum_{i=1}^n \left[(Y_i - \mu_{i+f_i})^2 + \lambda(f_i - f_{i-1})^2 \right],$$

where Y denotes a column of gel pixel values and μ a database column. However, to align many PFGE columns simultaneously on a single gel necessitates minimisation of

$$C(f, l) = \sum_{i=1}^n \sum_{j=1}^m \left[(Y_{ij} - \mu_{i+f_{ij}, l_j})^2 + \lambda_1(f_{ij} - f_{i-1, j})^2 + \lambda_2(f_{ij} - f_{i, j-1})^2 \right] \quad (2)$$

with respect to 2-D warping terms (f) and column labels (l), where Y is an $n \times m$ array of gel pixel values, μ is the 2-D array of database pixel values and (λ_1, λ_2) are non-negative constants. (Note, terms involving λ specify an intrinsic Gaussian Markov Random Field (MRF) for f with a first-order neighbourhood (Besag and Kooperberg, 1995).) Unfortunately, DP is unable to solve this 2-D problem if it involves more than three columns.

Image restoration or smoothing is another generic problem in image analysis. For example, synthetic aperture radar (SAR) is a form of remote sensing in which data have a large noise component due to speckle, and therefore need smoothing before manual or automatic interpretation (Oliver, 1991). Fig 2 shows a log-transformed C-band, HH-polarization, SAR image of an area near Thetford forest, England, in August 1989, obtained by plane as part of the Maestro-1 campaign (Joint Research Centre, Ispra, report IRSA/MWT/4.90), and previously used by Glasbey and Jones (1997) to evaluate nonlinear filtering algorithms. To restore an image we seek a 2-D array of values f , drawn again from a finite set, \mathcal{F} , that minimises

$$C(f) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left[(Y_{ij} - f_{ij})^2 + \phi\{f_{ij} - f_{i-1, j}\} + \phi\{f_{ij} - f_{i, j-1}\} \right], \quad (3)$$

where $\phi\{\cdot\}$ is a function to penalise changes in f , and again specifies an MRF with a first-order neighbourhood. If ϕ is convex then C can be minimised using standard gradient descent algorithms, except for the pathological case $\phi\{f\} = \lambda|f|$ (Künsch, 1994). However, it is desirable for ϕ to be non-convex in order to permit step changes in the restored values. Again, this is a 2-D problem for which DP is inapplicable, except in small cases. For example, Friel and Rue

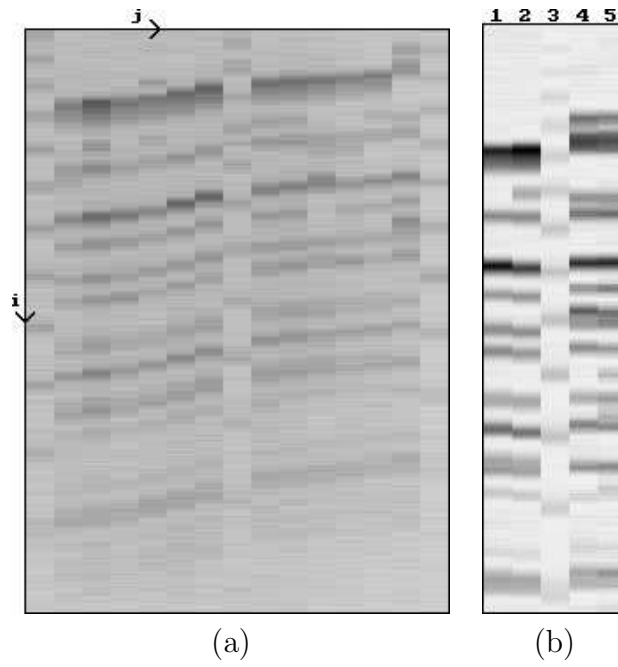


Figure 1: PFGE of *E. coli* O157 strains: **(a)** sample gel after preprocessing, **(b)** database of columns of known genotype.



Figure 2: Synthetic aperture radar (SAR) image of area near Thetford forest, England. Values range from 0, displayed as black, to 255, displayed as white.

(2007, section 3.2) proposed the use of DP to restore categorical images with 20 columns at most.

In §2 we present three generalisations of DP to 2-D. Then, in §3 we evaluate the algorithms for restoring the SAR image and warping PFGE gels. Finally, in §4 we discuss results and extensions.

2 Generalisations of DP

To simplify we will only consider minimising (3) in this section, and will address the modifications necessary to minimise (2) in §3.

2.1 Iterated dynamic programming (IDP)

The first generalisation of DP we consider is a greedy algorithm first proposed by Leung et al. (2004) for matching stereo pairs, termed iterated dynamic programming (IDP): DP is used to recursively solve each of a sequence of 1-D problems in turn, to find a local optimum. From an initial value of the 2-D array f , for each column j in turn, we minimise $C(f)$ w.r.t. $f_{:j} \equiv (f_{1j}, \dots, f_{nj})$, given current values of all other f 's, i.e.

$$f_{:j} = \arg \min_{(g_1, \dots, g_n) \in \mathcal{F}} \sum_{i=1}^n \left[(Y_{ij} - g_i)^2 + \phi\{g_i - g_{i-1}\} + \phi\{g_i - f_{i,j-1}\} + \phi\{g_i - f_{i,j+1}\} \right], \quad (4)$$

using (g_1, \dots, g_n) to represent possible values of (f_{1j}, \dots, f_{nj}) . We continue iterating through columns until C stops decreasing, and a local minimum has been found.

Before presenting technical details of how we solve (4), it may aid understanding to compare IDP with another greedy algorithm, ICM (Iterated Conditional Modes: Besag, 1986), which operates pixel-by-pixel rather than column-by-column. In ICM, for each pixel (i, j) in turn in a raster scan, $C(f)$ is minimised w.r.t. f_{ij} , given current values of all other f 's, i.e.

$$f_{ij} = \arg \min_{g \in \mathcal{F}} \left[(Y_{ij} - g)^2 + \phi\{g - f_{i-1,j}\} + \phi\{g - f_{i+1,j}\} + \phi\{g - f_{i,j-1}\} + \phi\{g - f_{i,j+1}\} \right],$$

using g to represent possible values of f_{ij} , and this is repeated until convergence. IDP can be viewed as a block-updating version of ICM. It requires of order $O(nm|\mathcal{F}|^2)$ calculations per iteration, whereas ICM requires of order $O(nm|\mathcal{F}|)$, where $|\mathcal{F}|$ denotes the number of elements in \mathcal{F} . However, by optimising a whole column at once we would hope that IDP either converges in fewer iterations or to a lower value of C .

Returning to IDP, we can solve (4) using DP, which involves a forward recursion followed by a backward recursion. In the forward recursion, two arrays, D and H , are computed as follows:

$$D_{1g} = \left[(Y_{1j} - g)^2 + \phi\{g - f_{1,j-1}\} + \phi\{g - f_{1,j+1}\} \right] \quad \forall g \in \mathcal{F},$$

and recursively for $i = 2, \dots, n$,

$$H_{ig} = \arg \min_{h \in \mathcal{F}} W_h \quad D_{ig} = \min_{h \in \mathcal{F}} W_h \equiv W_{H_{ig}} \quad \forall g \in \mathcal{F},$$

where $W_h = D_{i-1,h} + [(Y_{ij} - g)^2 + \phi\{g - h\} + \phi\{g - f_{i,j-1}\} + \phi\{g - f_{i,j+1}\}]$,

using (h, g) to represent possible values of $(f_{i-1,j}, f_{ij})$. In column j , conditional on the restoration taking value g in row i , H_{ig} is the optimal restoration for row $(i - 1)$ and D_{ig} is the cumulative cost of the optimal restoration for rows 1 to i . Following the forward recursion, in the backward recursion optimal values of $f_{.j}$ are obtained simply by

$$f_{nj} = \arg \min_{g \in \mathcal{F}} D_{ng} \quad \text{and} \quad f_{ij} = H_{i+1, f_{i+1,j}} \quad \text{recursively for } i = (n - 1), \dots, 1.$$

2.2 Stochastic IDP

A drawback of IDP, which ICM shares, is that they are not global optimisers, but become trapped in local optima of C . One way to partially alleviate this problem is to make the IDP algorithm stochastic. Hence, a second algorithm replaces C in IDP by

$$C(f) + \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m e_{ij, f_{ij}}, \quad \text{where } e_{ijf} \sim U[0, T] \quad \forall i, j, f,$$

and $U[0, T]$ denotes a uniform distribution over the interval 0 to T . Thus, the value of array f that minimises this expression follows a probability distribution determined by T , albeit one that we cannot express analytically. The temperature, T , is initially set at a sufficiently large value, so that all values of f are equally likely. We follow an exponential cooling schedule: after each iteration, $T \searrow \alpha T$, where α is a constant less than unity, and e is resampled. The algorithm is run until convergence when T is sufficiently small that e is negligible.

2.3 Forward-backwards simulated annealing (FB-SA)

Stochastic IDP is an *ad hoc* modification to IDP and does not guarantee finding the global optimum, unlike simulated annealing, which does provided that we sample f proportional to $\exp[-C(f)/T]$ and reduce T sufficiently slowly (Kirkpatrick et al., 1983; Geman and Geman, 1984). Pixel-by-pixel Gibbs sampling can be used to simulate a Markov chain that has this limit distribution: for each pixel (i, j) in turn in a raster scan, f_{ij} is sampled proportional to

$$\exp \left[-\frac{1}{T} \left((Y_{ij} - f_{ij})^2 + \phi\{f_{ij} - f_{i-1,j}\} + \phi\{f_{ij} - f_{i+1,j}\} + \phi\{f_{ij} - f_{i,j-1}\} + \phi\{f_{ij} - f_{i,j+1}\} \right) \right],$$

conditional on current values of all other f 's. As in §2.2, the temperature, T , is initially set at a sufficiently large value, so that all values of f are equally likely. Again we follow an exponential cooling schedule: after each iteration, $T \searrow \alpha T$, and the algorithm is run until convergence. (Note, this is a standard cooling schedule although it is too fast to guarantee finding the global optimum.) We subsequently refer to this algorithm as SA.

Rather than sample pixel-by-pixel we can sample column-by-column as in §2.2, using a Forward-Backwards Gibbs Sampler (FB), first proposed by Eddy (1995) and subsequently named by Scott (2002). This is a form of block-updating Gibbs sampler (Roberts and Sahu, 1997) with each column a block. For each column j in turn, we sample $f_{.j}$ with probability proportional to $e^{-C(f)/T}$, given current values of all other f 's, i.e. sample proportional to

$$\exp \left[-\frac{1}{T} \sum_{i=1}^n \left((Y_{ij} - f_{ij})^2 + \phi\{f_{ij} - f_{i-1,j}\} + \phi\{f_{ij} - f_{i,j-1}\} + \phi\{f_{ij} - f_{i,j+1}\} \right) \right]. \quad (5)$$

After each iteration, $T \searrow \alpha T$, and the algorithm is run until convergence. We term this algorithm forward-backwards simulated annealing (FB-SA).

We sample (5) for column j , using the FB algorithm. This has similarities with DP, in that the first step is a forward recursion, but this is then followed by a backward stochastic sampler, rather than another deterministic recursion. In the forward recursion an array, D , is computed as follows:

$$D_{1g} = \exp \left[-\frac{1}{T} \left((Y_{1j} - g)^2 + \phi\{g - f_{1,j-1}\} + \phi\{g - f_{1,j+1}\} \right) \right] \quad \forall g \in \mathcal{F},$$

and recursively for $i = 2, \dots, n$,

$$D_{ig} = \sum_{h \in \mathcal{F}} D_{i-1,h} \exp \left[-\frac{1}{T} \left((Y_{ij} - g)^2 + \phi\{g - h\} + \phi\{g - f_{i,j-1}\} + \phi\{g - f_{i,j+1}\} \right) \right] \quad \forall g \in \mathcal{F}.$$

D_{ig} is proportional to the marginal probability of the restoration taking value g in row i , given information in rows 1 to i . In the backward stochastic sampler, values of $f_{.j}$ are obtained by

$$f_{nj} = g \quad \text{with probability} \quad \frac{D_{ng}}{\sum_h D_{nh}} \quad \forall g \in \mathcal{F},$$

and recursively for the decreasing sequence, $i = (n-1), \dots, 1$,

$$f_{ij} = g \quad \text{with probability} \quad \frac{D_{ig} \exp \left[-\frac{1}{T} \phi\{g - f_{i+1,j}\} \right]}{\sum_h D_{ih} \exp \left[-\frac{1}{T} \phi\{h - f_{i+1,j}\} \right]} \quad \forall g \in \mathcal{F}.$$

Computational implementation of these steps needs care to avoid underflow errors as $T \searrow 0$.

3 Results

The three generalisations of DP presented in §2, together with the two pixel-by-pixel methods, ICM and SA, were used to restore the SAR image in Fig 2, by minimising (3). Because values of Y are in the range 0 to 255, and values near the extremes are relatively rare, we restricted to a set of values for f of $\mathcal{F} = \{65, 75, \dots, 165\}$, and for a simple, non-convex, smoothness penalty, we used

$$\phi\{z\} = \begin{cases} 0 & \text{if } z = 0 \\ \lambda & \text{otherwise,} \end{cases} \quad (6)$$

with the penalty value of $\lambda = 2000$ chosen by eye to produce realistic restorations. When the IDP algorithm was used, starting from $f = Y$, the restoration had a clear artefact due to repeated smoothing down columns. By modifying the algorithm to alternate between columns and rows this effect was reduced. However, it is still evident in Fig 3(a), which shows the resulting restoration, with a minimised value of $C = 515$. Fig 3(b) shows the restoration produced by ICM, also starting from $f = Y$, resulting in $C = 885$. Again, a directional artefact is evident in the restoration, although this has been lessened by again alternating between raster scans along rows and down columns. In this case, pixel values vary less in a diagonal direction from top-left to bottom-right.

Because IDP and ICM are deterministic algorithms, they are prone to get trapped in local optima, and lower values of C can be obtained using different starting values for f . We initialised f with outputs of moving median filters of various sizes (see, for example, Glasbey and Horgan, 1995, pp. 78-82). The optimal window size for IDP was 9×9 , resulting in the lowest value found for C of 466 and the restoration shown in Fig 3(c), whereas for ICM it was 35×35 , resulting in $C = 553$ and the restoration shown in Fig 3(d), although to the human eye this looks oversmoothed, and illustrates the problem with local optima.

In the stochastic algorithms, the rate of cooling parameter, α , was chosen such that the total number of iterations was approximately $2^8, 2^9 \dots$. For both FB-SA and SA, T was initialised at 10^5 and convergence was reached around 10^{-3} , whereas for stoch-IDP, corresponding values were 10^7 and 10^{-1} . So, for all algorithms, if the target number of iterations is 2^8 , $\alpha = 10^{-8/2^8} = 0.93057$. Fig 4(a) shows minimised values of C for runs of all the algorithms, plotted against the number of iterations on a log-scale. We see that all three stochastic algorithms found lower values of C than the two deterministic algorithms provided the cooling schedule was slow enough. Although the stochastic IDP algorithm has the least theoretical basis, it takes fewer iterations than FB-SA to achieve a particular value of C , which in turn outperforms SA. However, when we consider CPU time, plotted in Fig 4(b) on a log-scale, it appears that FB-SA and SA take similar times to achieve particular values of C , although stoch-IDP remains best. The lowest value of C was found by stoch-IDP (438), and the corresponding restoration is shown in Fig 3(e), although we would expect the other stochastic algorithms to produce similar results given a slower cooling schedule and more CPU time.

Pixel-by-pixel algorithms perform poorly for warping problems, so we only applied the three column-based algorithms to minimise (2) for the PFGE data. We used $\mathcal{F} = \{0, 0.2, 0.4, 0.6, \dots\}$ and constrained $(f_i - f_{i-1}) = \{0.2, 0.4, \dots, 1.8\}$. We also set $\lambda = (10, 1)$, the pair of values identified by Gustafsson (2005) using column-wise cross-validation. Fig 5 shows the minimised values of C plotted against CPU time for multiple runs of the stochastic algorithms applied to the data in Fig 1(a) and to a second gel. We started IDP from $f = 0$, but we also show results of a modification, which we term DP-IDP, where initially DP is applied separately in each column:

$$(f_{.j}, l_j) = \arg \min_{(g_1, \dots, g_{n,p})} \sum_{i=1}^n [(Y_{ij} - \mu_{i+g_i,p})^2 + \lambda_1 (g_i - g_{i-1})^2].$$

We see that this deterministic algorithm performs best. Stochastic IDP found the same optimal solution for gel 1, provided the cooling rate was not too rapid, but both stochastic algorithms appear to become trapped in local optima for gel 2. Fig 6 shows the results of warping the two

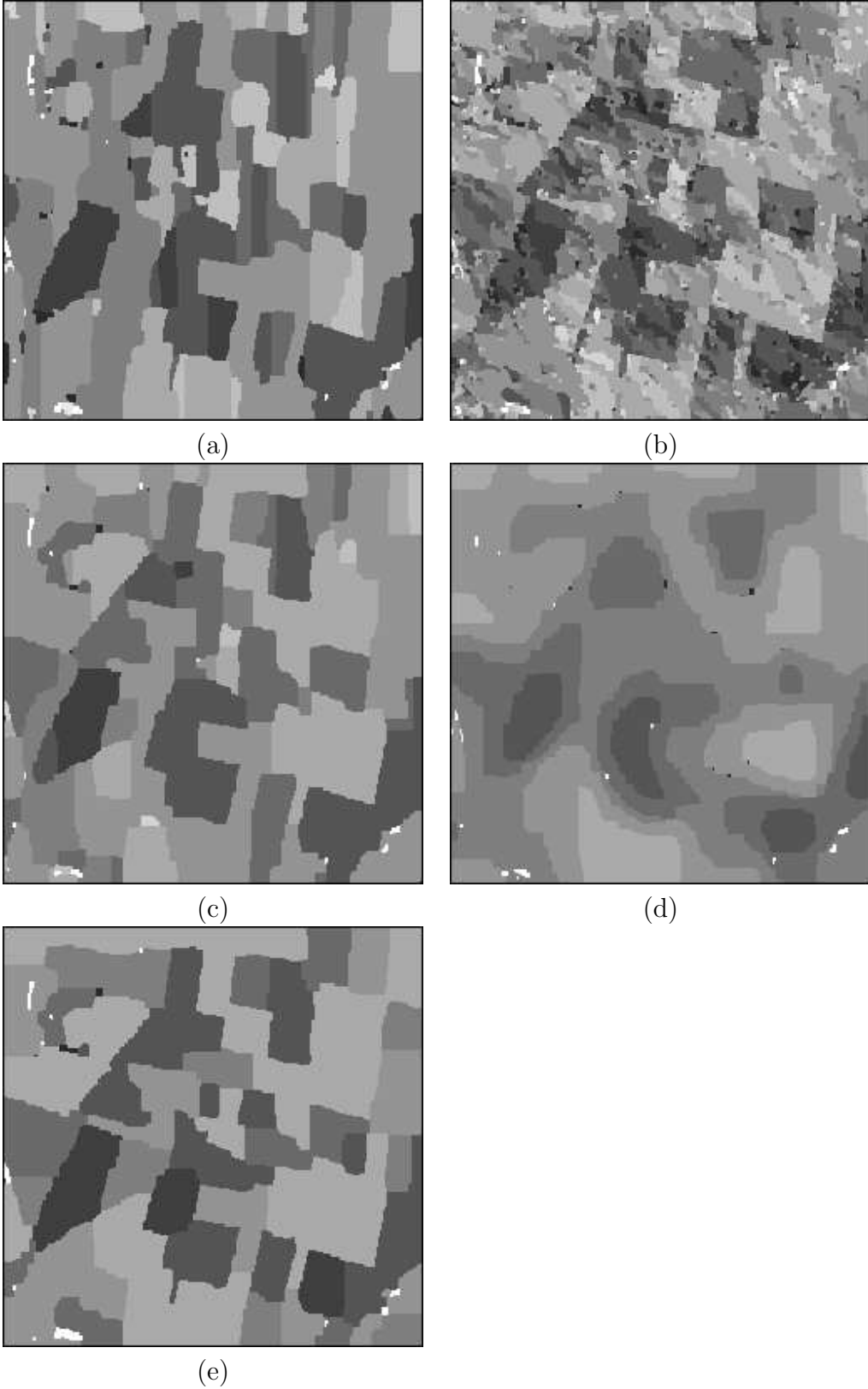


Figure 3: Restorations of SAR image in Fig 2 using different algorithms: **(a)** IDP from $f = Y$ ($C = 515$); **(b)** ICM from $f = Y$ ($C = 885$); **(c)** IDP from $f = 9 \times 9$ moving median filter of Y ($C = 466$); **(d)** ICM from $f = 35 \times 35$ moving median filter of Y ($C = 553$); **(e)** best result from stochastic IDP ($C = 438$).

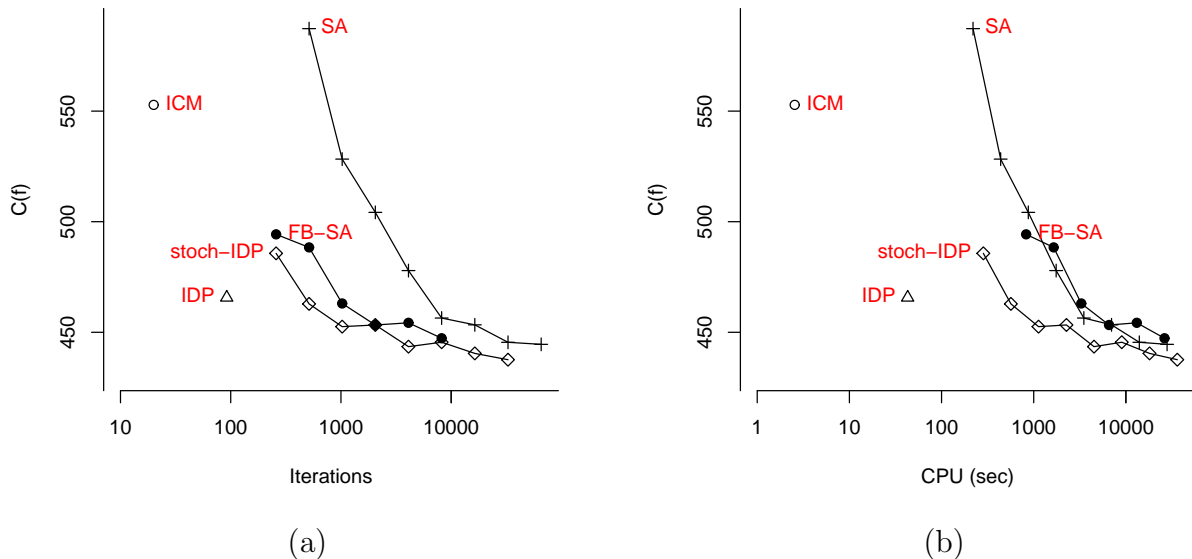


Figure 4: Minimised values of C , for runs of algorithms to restore SAR image, plotted against (a) number of iterations and (b) CPU time using Fortran90 on a SunUltra5.

gels using the estimated values of \hat{f} from ID-IDP, and the estimated column labels \hat{l} .

4 Discussion

We have considered some 2-D generalisations of DP, in order to solve generic problems in image analysis. One generalisation is deterministic, IDP, and two are stochastic, stoch-IDP and FB-SA. We contrasted these algorithms with two pixel based ones, ICM, which is deterministic, and SA, which is stochastic. In several applications, including those in §3, IDP performed well, particularly if care was taken with the starting value for array f . Of the stochastic methods, stoch-IDP proved to be best, in spite of its lack of a theoretical basis. At the price of considerably greater CPU times, it produced a superior SAR restoration, though it failed to improve on the PFGE warping found by IDP, which is possibly a more challenging problem because of the need to also identify column labels. Therefore, our recommendation is to use of IDP from multiple starts, following which it is relatively straightforward to extend computer code to stoch-IDP if there are problems with local optima.

The algorithms are open to many extensions and modifications, some of which we have already seen in §3. All our examples involved MRFs with a first-order neighbourhood, but extend to higher-orders, albeit at greater CPU and memory costs. Although IDP, stoch-IDP and FB-SA were all presented as column-by-column methods in §2, DP and FB could be used to find f for any non-intersecting path through the $n \times m$ array. Also, although it is not possible to apply either DP or FB to a full 2-D array, it would be possible to consider two or possibly

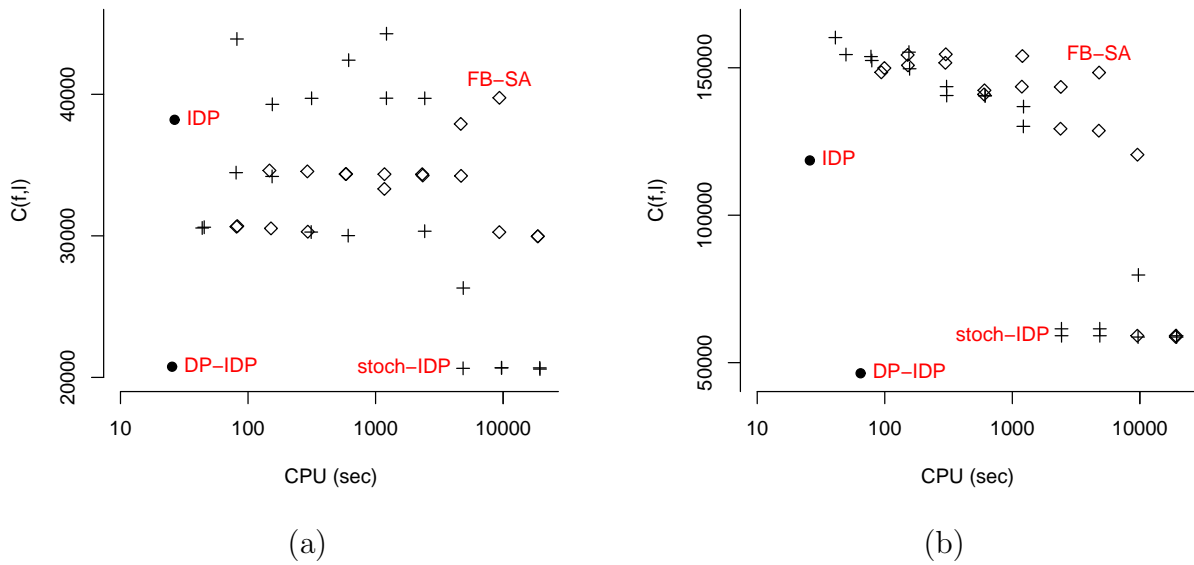


Figure 5: Minimised values of C plotted against CPU time, using Fortran90 on a SunUltra5, for runs of algorithms to warp: (a) PFGE gel shown in Fig 1(a), (b) a second gel.

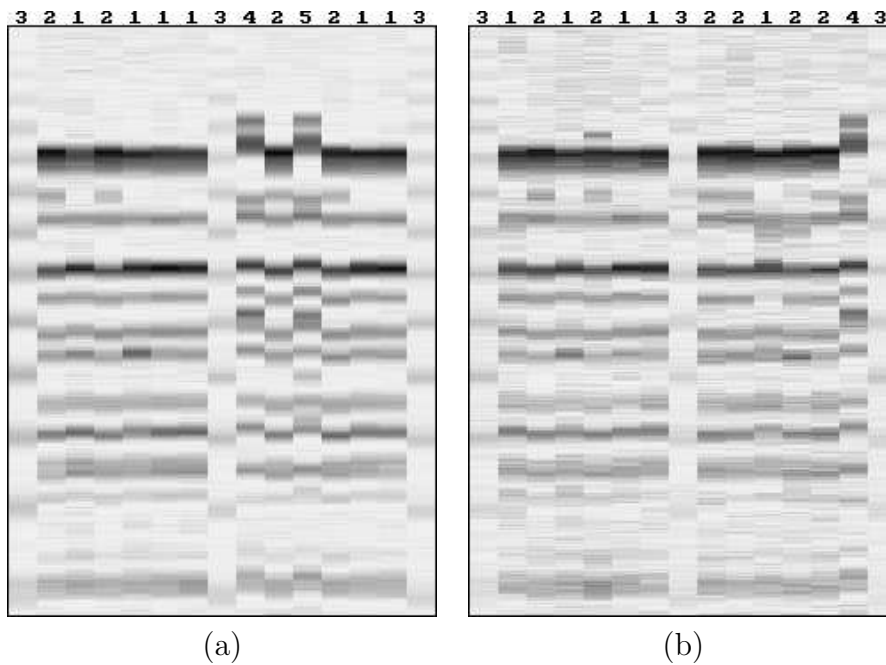


Figure 6: Warped gels, labelled with most similar columns in database, obtained using ID-IDP: (a) PFGE gel shown in Fig 1(a), (b) a second gel.

three columns simultaneously. Further, the algorithms may run faster if we restrict to a coarser subset of \mathcal{F} , and a coarser grid of pixels, during early iterations, and thus adopt a multiresolution approach, or only search local values for f in later iterations and neighbourhoods of changed values from the previous iteration. There may also be gains from varying λ during the course of the algorithms. It remains to be determined whether any of these modifications is beneficial, the evidence from §3 being that conclusions are likely to be application specific.

It would be interesting to know whether there is any cooling schedule for which stoch-IDP could be shown to guarantee a global optimum, even though it may be as impractically slow as the standard one for simulated annealing. We assumed λ known in both applications. If instead we wanted to estimate it, we could formulate a Bayesian model and use the FB-SA algorithm, but without the cooling schedule, to sample from the posterior distributions. A broader comparison could also be conducted with other energy minimisation methods such as Lagrange relaxation (Storvik and Dahl, 2000), graph cuts (Boykov et al., 2001), belief propagation and message passing (Szeliski et al., 2006). A graph cut algorithm has previously been used by Greig et al. (1989) to obtain exact MAP estimates to restore binary images.

The generalisations of DP have broad applicability. As with matching of stereo pairs, the PFGE problem can be regarded as $1\frac{1}{2}$ -D warping because columns remain fixed, but the methods can be extended to full 2-D warping, for example to align SDS-PAGE electrophoresis gels. We have also used IDP successfully to segment 3-D images, by extending (1) to find the 2-D surface, f , that minimises

$$C(f) = \sum_{i=1}^n \sum_{j=1}^m \left[Y_{i,j,f_{ij}}^* + \lambda_1 (f_{i,j} - f_{i-1,j})^2 + \lambda_2 (f_{i,j} - f_{i,j-1})^2 \right],$$

where Y_{ijk}^* denotes an element in a 3-D array, the cost of the surface passing through 3-D location (i, j, k) (Navajas et al., 2006). Further, image warping has similarities with alignment of multiple DNA sequences, which can be formulated as finding f to minimise

$$C(f) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^m \phi \left\{ Y_{i+f_{ij},j} - Y_{i+f_{ik},k} \right\} + \sum_{i=1}^n \sum_{j=1}^m (f_{i,j} - f_{i-1,j})^2,$$

where Y_j denotes the j th DNA sequence and ϕ is defined in (6). Here, f is not constrained to be smooth between columns, but all data have to be warped to a consensus sequence. DP can be used to align pairs of sequences, but not more. This was the original application for the Forward-Backwards Gibbs Sampler (Eddy, 1995), and for alternative approaches, see (Notredame, 2002), but the remaining algorithms in §2 could also be adapted for this problem.

Acknowledgements

I am grateful to my colleague, Dirk Husmeier, for bringing to my attention the Forward-Backwards Gibbs Sampler. The work was funded by the Scottish Executive.

References

- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton.
- Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B*, 48:259–302.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.
- Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222–1239.
- Eddy, S. R. (1995). Multiple alignment using hidden Markov models. In Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T., and Wodak, S., editors, *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*, pages 114–120, Menlo Park, California. The AAAI Press.
- Friel, N. and Rue, H. (2007). Recursive computing and simulation-free inference for general factorizable models. *Biometrika*, 94:661–672.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–735.
- Glasbey, C. A. and Horgan, G. W. (1995). *Image Analysis for the Biological Sciences*. Wiley, Chichester.
- Glasbey, C. A. and Jones, R. (1997). Fast computation of moving average and related filters in octagonal windows. *Pattern Recognition Letters*, 18:555–565.
- Glasbey, C. A., Vali, L., and Gustafsson, J. S. (2005). A statistical model for unwarping of 1-D electrophoresis gels. *Electrophoresis*, 26:4237–4242.
- Glasbey, C. A. and Young, M. J. (2002). Maximum *a posteriori* estimation of image boundaries by dynamic programming. *Applied Statistics*, 51:209–221.
- Greig, D. M., Porteous, B. T., and Seheult, A. H. (1989). Exact maximum *a posteriori* estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51:271–279.
- Gustafsson, J. (2005). *Unwarping and Analysing Electrophoresis Gels*. PhD thesis, Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220:671–680.
- Künsch, H. R. (1994). Robust priors for smoothing and image restoration. *Annals of the Institute of Statistical Mathematics*, 46:1–19.
- Leung, C., Appleton, B., and Sun, C. (2004). Fast stereo matching by Iterated Dynamic Programming and quadtree subregioning. In Hoppe, A., Barman, S., and Ellis, T., editors, *British Machine Vision Conference*, volume 1, pages 97–106.

- MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London.
- Navajas, E. A., Glasbey, C. A., McLean, K. A., Fisher, A. V., Charteris, A. J. L., Lambe, N. R., Bunger, L., and Simm, G. (2006). In vivo measurements of muscle volume by automatic image analysis of spiral computed tomography scans. *Animal Science*, 82:545–553.
- Notredame, C. (2002). Recent progress in multiple sequence alignment: a survey. *Pharmacogenomics*, 3:131–144.
- Oliver, C. J. (1991). Information from SAR images. *Journal of Physics D: Applied Physics*, 24:1493–1514.
- Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society, Series B*, 59:291–317.
- Scott, S. L. (2002). Bayesian methods for Hidden Markov Models: recursive computing in the 21st century. *Journal of the American Statistical Association*, 97:337–351.
- Storvik, G. and Dahl, G. (2000). Lagrangian-based methods for finding MAP solutions for MRF models. *IEEE Transactions on Image Processing*, 9:469–479.
- Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2006). A comparative study of energy minimization methods for Markov random fields. In *Computer Vision – ECCV, Part 2*, number 3952 in Lecture Notes in Computer Science, pages 16–29. Springer, Berlin.