Registering Incomplete Radar Images using the EM Algorithm

Simon Moss and Edwin R. Hancock
Department of Computer Science, University of York
York, Y01 5DD, UK

Abstract

This paper describes an application of the EM (expectation and maximisation) algorithm to the registration of incomplete millimetric radar images. The data used in this study consists of a series of non overlapping radar sweeps. Our registration process aims to recover transformation parameters between the radar-data and a digital map. The tokens used in the matching process are fragmented line-segments extracted from the radar images which predominantly correspond to hedge-rows in the cartographic data. The EM technique models data uncertainty using Gaussian mixtures defined over the positions and orientations of the lines. The resulting weighted least-squares parameter estimation problem is solved using the Levenberg-Marquardt method. A sensitivity analysis reveals that the data-likelihood function is unimodal in the translation and scale parameters. In-fact the algorithm is only potentially sensitive to the choice of initial rotation parameter; this is attributable to local sub-optima in the log-likelihood function associated with $\frac{\pi}{2}$ orientation ambiguities in the map.

1 Introduction

Radar imagery is an important means of sensing and mapping elevated features in the landscape. Unfortunately, one of the obstacles to the automatic interpretation of this imagery is the presence of speckle noise. The speckle process is most intrusive when the scene contains physical structure whose spatial scale is of the same order as the wavelength of the radar. In a recent study, we demonstrated how relaxation techniques could be successfully deployed to control clutter in the segmentation and matching of features in conventional synthetic aperture radar images [4, 12]. The work reported in this paper describes a more ambitious programme of research aimed at matching millimetric Doppler beam sharpened images (DBS) against a digital map. These images differ from their SAR counterparts in a number of important respects. Firstly, the frequency of the radar is of the order of 100GHz rather than the 10 GHz which is typical of SAR. This means that structures whose size is of the order of a few millimetres appear rough to the radar. The shorter wavelengths employed in the DBS imagery are a consequence of physical constraints imposed upon the dimensions of the resonating cavities in airborne military radars. The second difficulty stems from the imaging geometry.

BMVC 1996 doi:10.5244/C.10.34

Since the radar is used to sense objects in the line of flight from a low flying aircraft, the images are subject to small angle systematics. Finally, the focussing of the radar means that the scenes are imaged as a series of non-overlapping sweeps interspersed with dead-regions.

The practical goal of the work described in this paper is to develop a navigation aid for airborne vehicles. The scenes under study are of rural areas where the principal man-made features available for cartographic matching are linear hedge structures. These are elevated and produce ridge artifacts in the radar images. Since these features are vegetative, they appear rough to the radar and result in specularities. The main obstacle to cartographic matching resides in the fact that because the scene is sensed as a series of non-overlapping sweeps the radar fragments the target hedge structures. This problem is illustrated in the example image shown in Figure 1a. Preliminary studies have shown that because we are operating both with highly fragmented line-tokens and severe levels of clutter, conventional grouping strategies provide an ineffective means of preprocessing prior to matching [12]. Instead, we seek a statistical framework that is capable of recovering the transformation between radar-data and map when clutter and data incompleteness are limiting factors.

It is for these reasons that we turn to Dempster, Laird and Rubin's EM (expectation and maximisation) [3] algorithm. This algorithm was originally introduced as a means of finding maximum likelihood solutions to problems posed in terms of incomplete data. Despite its relatively poor convergence properties, the algorithm provides a powerful statistical framework for fitting sparse data that has many features in common with robust parameter estimation. In the domain of machine vision, the EM algorithm has been exploited for estimating multiple motion parameters [9] and for face recognition [10]. Of particular relevance to the work reported in this paper, is the fact that the algorithm has been successfully utilised in the recognition of occluded objects [13] and in the extraction of 3D object pose [6] from relatively uncluttered 2D images. Viola and Wells [13] have reported a registration method which has many features that are reminiscent of the EM algorithm. The method maximises a mutual information measure defined over a set of Parzen estimates to align images using raw intensity information. Recently, the methodological basis of the algorithm has attracted renewed interest in the domain of artificial neural networks where it has not only been shown to have an intimate relationship with mean-field annealing [1], but also to provide a convenient framework for hierarchical data processing [7]. The basic idea underlying the algorithm is to iterate between the expectation and maximisation steps until convergence is reached. Expectation involves computing a weighted likelihood function using a mixture density specified in terms of a series of model parameters. In the maximisation phase, the model parameters are recomputed to maximise the expected value of the incomplete data likelihood.

2 The EM Algorithm

In this Section we detail our representation of the matching process and describe how the underlying set of transformation parameters can be recovered using the EM algorithm. The EM algorithm was first introduced by Dempster, Laird and Rubin as a means of fitting incomplete data [3]. The algorithm has two stages.

The expectation step involves estimating a mixture distribution using current parameter values. The maximisation step involves computing new parameter values that optimise the expected value of the weighted data likelihood. This two-stage process is iterated to convergence. Although the EM algorithm has been exploited in the recovery of object pose by Hornegger and Nieman [6], the main contribution of the this paper is to demonstrate the effectiveness of the algorithm in matching the highly cluttered and incomplete imagery delivered by millimetric radar.

2.1 Representation

Our basic aim is to recover the parameters of the co-ordinate transformation between the incomplete radar data and a digital map. The tokens used in the matching process are line-segments which are characterised by their mid-point co-ordinates (x_i, y_i) in the image plane and their line-orientation in the image co-ordinate system θ_i. Each line in the radar-image is represented by a vector $\underline{\mathrm{w}}_i = (x_i, y_i, \theta_i)^T$ where i is the segment index. The available line-data for the radar image is denoted by the set $\mathbf{w} = \{\underline{\mathrm{w}}_i, \forall i \in \mathcal{D}\}$ where $\mathcal{D}$ is the segment index-set. The lines constituting the cartographic model are similarly represented by the set $\mathbf{z} = \{\underline{\mathrm{z}}_j, \forall j \in \mathcal{M}\}$. Here $\mathcal{M}$ is the index-set for the model lines and the $\underline{\mathrm{z}}_j$ represent the corresponding measurement-vectors. The aim of our matching algorithm is to iteratively recover a parameter-vector $\Phi^{(n)} = (\phi_1^{(n)}, ..., \phi_4^{(n)})^T$ which describes the Euclidean transformation that brings the radar lines and map lines into registration with one-another. The four components of the parameter vector are as follows; $\phi_1^{(n)}$ represents the x-translation, $\phi_2^{(n)}$ represents the y-translation, $\phi_3^{(n)}$ is the rotation and $\phi_4^{(n)}$ is the relative scale. The registration process is effected by transforming the measurement vectors representing the model-lines in the map into the co-ordinate system of the radar-image. The transformed version of the measurement vector $\underline{\mathrm{z}}_j$ is given by

$$\underline{\mathrm{z}}'_j = F(\underline{\mathrm{z}}_j, \Phi^{(n)}) \tag{1}$$

where the transformation-function F is defined as follows

$$F(\underline{\mathrm{z}}_j, \Phi^{(n)}) = U(\Phi^{(n)})\underline{\mathrm{z}}_j + V\Phi^{(n)} \tag{2}$$

The matrix $U(\Phi^{(n)})$ models the scaling and rotation of co-ordinates

$$U(\Phi^{(n)}) = \begin{pmatrix} \phi_4^{(n)} \cos\phi_3^{(n)} & -\phi_4^{(n)} \sin\phi_3^{(n)} & 0 \\ \phi_4^{(n)} \sin\phi_3^{(n)} & \phi_4^{(n)} \cos\phi_3^{(n)} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

while the matrix V selects the translation components for the parameter vector $\Phi^{(n)}$

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{4}$$

2.2 Expectation

Basic to our philosophy of exploiting the EM algorithm is the idea that every line-segment in the radar data can in principle associate to each of the lines in the digital map with some *a posteriori* probability. This modelling ingredient is naturally incorporated into the fitting process by developing a mixture model over the space of potential matching assignments. The expectation step of the EM algorithm provides an iterative framework for computing the *a posteriori* matching probabilities using Gaussian mixtures defined over a set of transformation parameters.

The EM algorithm commences by considering the conditional likelihood for the data-line measurements $\underline{\mathrm{w}}_i$ given the current set of transformation parameters, $\Phi^{(n)}$. The algorithm builds on the assumption that the individual data items are conditionally independent of one-another given the current parameter estimates, i.e.

$$p(\mathbf{w}|\Phi^{(n)}) = \prod_{i \in \mathcal{D}} p(\underline{\mathrm{w}}_i|\Phi^{(n)}) \tag{5}$$

Each of the component densities appearing in the above factorisation is represented by a mixture distribution defined over a set of putative model-data associations between the two sets of line-tokens

$$p(\underline{\mathrm{w}}_i|\Phi^{(n)}) = \sum_{j \in \mathcal{M}} p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n)}) P(\underline{\mathrm{z}}_j|\Phi^{(n)}) \tag{6}$$

The ingredients of the above mixture density are the component conditional measurement densities $p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n)})$ and the mixing proportions $P(\underline{\mathrm{z}}_j|\Phi^{(n)})$. The conditional measurement densities represent the likelihood that the data-line measurement $\underline{\mathrm{w}}_i$ originates from the model-line indexed j under the prevailing set of transformation parameters $\Phi^{(n)}$. We use the shorthand notation $\alpha_j^{(n)} = P(\underline{\mathrm{z}}_j|\Phi^{(n)})$ to denote the mixing proportions. These quantities provide a natural mechanism for assessing the significance of the individual model-lines in explaining the current data-likelihood. For instance if $\alpha_j^{(n)}$ approaches zero, then this indicates that there is no matching line in the data. In other words, the mixture model provides a natural way of accommodating missing or occluded model segments. It is important to stress that the mixing proportions are iteration dependent, being conditioned upon the current parameter values.

Conventionally, maximum-likelihood parameters are estimated using the complete log-likelihood for the available data

$$L(\Phi^{(n)}, \mathbf{w}) = \sum_{i \in \mathcal{D}} \ln p(\underline{\mathrm{w}}_i|\Phi^{(n)}) \tag{7}$$

In the case where the conditional measurement densities are univariate Gaussian, then maximising the complete likelihood function corresponds to solving a system of least-squares equations for the transformation parameters. By contrast, the expectation step of the EM algorithm is aimed at estimating the log-likelihood function when the data under consideration is incomplete. In our line-matching example this incompleteness is a consequence of the fact that we do not know how to associate tokens in the radar image and their counterparts in the cartographic

map. In other words we need to average the log-likelihood over the space of potential line associations. In fact, it was Dempster, Laird and Rubin [3] who observed that maximising the weighted log-likelihood was equivalent to maximising the conditional expectation of the log-likelihood for a new parameter set given an old parameter set. For our matching problem, maximisation of the expectation of the conditional likelihood, i.e. $E[L(\Phi^{(n+1)}, \mathbf{w})|\Phi^{(n)}, \mathbf{w})]$, is equivalent to maximising the weighted log-likelihood function

$$Q(\Phi^{(n+1)}|\Phi^{(n)}) = \sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{M}} P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)}) \ln p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n+1)}) \tag{8}$$

The *a posteriori* probabilities $P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)})$ play the role of matching weights in the expected likelihood. We interpret these weights as representing the probability of match between the data line indexed i and the model-line indexed j. In other words, they represent model-datum affinities. Using the Bayes rule, we can rewrite the *a posteriori* matching probabilities in terms of the components of the conditional measurement densities appearing in the mixture model in equation (6)

$$P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)}) = \frac{\alpha_j^{(n)} p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n)})}{\sum_{j' \in \mathcal{M}} \alpha_{j'}^{(n)} p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_{j'}, \Phi^{(n)})} \tag{9}$$

The mixing proportions are computed by averaging the *a posteriori* probabilities over the set of data-lines, i.e.

$$\alpha_j^{(n+1)} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)}) \tag{10}$$

In order to proceed with the development of a line registration process we require a model for the conditional measurement densities, i.e. $p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n)})$. Here we assume that the required model can be specified in terms of a multivariate Gaussian distribution. The random variables appearing in these distributions are the error residuals for the position and orientation predictions of the jth model line delivered by the current estimated transformation parameters. Accordingly we write

$$p(\underline{\mathrm{w}}_i|\underline{\mathrm{z}}_j, \Phi^{(n)}) = \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}\epsilon_{i,j}(\Phi^{(n)})^T \Sigma^{-1} \epsilon_{i,j}(\Phi^{(n)})\right] \tag{11}$$

In the above expression Σ is the variance-covariance matrix for the vector of error-residuals $\epsilon_{i,j}(\Phi^{(n)}) = \underline{\mathrm{w}}_i - F(\underline{\mathrm{z}}_j, \Phi^{(n)})$ between the components of the predicted measurement vectors $\underline{\mathrm{z}}'_j$ and their counterparts in the data, i.e. $\underline{\mathrm{w}}_i$. Formally, the matrix is related to the expectation of the outer-product of the error-residuals i.e. $\Sigma = E[\epsilon_{i,j}(\Phi^{(n)})\epsilon_{i,j}(\Phi^{(n)})^T]$. Accordingly, we compute the following estimate of Σ,

$$\tilde{\Sigma} = \frac{\sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{M}} P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)}) \epsilon_{i,j}(\Phi^{(n)}) \epsilon_{i,j}(\Phi^{(n)})^T}{\sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{M}} P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)})} \tag{12}$$

With these ingredients, and using the shorthand notation $q_{i,j}^{(n)} = P(\underline{\mathrm{z}}_j|\underline{\mathrm{w}}_i, \Phi^{(n)})$

for the *a posteriori* matching probabilities, the expectation step of the EM algorithm simply reduces to computing the weighted squared error criterion

$$Q'(\Phi^{(n+1)}|\Phi^{(n)}) = -\frac{1}{2}\sum_{i\in\mathcal{D}}\sum_{j\in\mathcal{M}} q_{i,j}^{(n)}\epsilon_{i,j}(\Phi^{(n)})^T\tilde{\Sigma}^{-1}\epsilon_{i,j}(\Phi^{(n)}) \tag{13}$$

In other words, the *a posteriori* probabilities $q_{i,j}^{(n)}$ effectively regulate the contributions to the likelihood function. Matches for which there is little evidence contribute insignificantly, while those which are in good registration dominate.

2.3 Maximisation

The maximisation step aims to locate the updated the parameter-vector $\Phi^{(n+1)}$ that optimises the quantity $Q(\Phi^{(n+1)}|\Phi^{(n)})$, i.e.

$$\Phi^{(n+1)} = \arg\max_{\Phi} Q'(\Phi|\Phi^{(n)}) \tag{14}$$

We solve the implied weighted least-squares minimisation problem using the Levenberg Marquardt technique [8, 11]. This non-linear optimisation technique offers a compromise between the steepest gradient and inverse Hessian methods. The former is used when close to the optimum while the latter is used far from it. In other words, when close to the optimum, parameter updating takes place with step-size proportional to the gradient $\nabla_\Phi Q'(\Phi|\Phi^{(n)})$. When far from the optimum the optimisation procedure uses second-order information residing in the Hessian, H, of $Q'(\Phi|\Phi^{(n)})$; the corresponding step-size for the parameter vector Φ is $H^{-1}\nabla_\Phi Q(\Phi|\Phi^{(n)})$. Central to the Levenberg-Marquardt method is the idea of exerting control over these two update modes using a positive parameter λ. This parameter defines the elements of the matrix Λ

$$\Lambda_{k,l} = \begin{cases} 1+\lambda & \text{if } k = l \\ 1 & \text{otherwise} \end{cases} \tag{15}$$

According to the Levenberg-Marquardt method the step-size $\delta\phi_l$ for the parameter ϕ_l is found by solving the following set of linear equations

$$\sum_{l=1}^{4}\Lambda_{k,l}\frac{\partial^2 Q'(\Phi|\Phi^{(n)})}{\partial\phi_k\partial\phi_l}\delta\phi_l = \frac{\partial Q'(\Phi|\Phi^{(n)})}{\partial\phi_k} \tag{16}$$

The parameter λ is chosen to be large if $\nabla_\Phi Q'(\Phi|\Phi^{(n)}) > 0$; in this case the optimisation process operates in steepest gradient mode. If on the other hand, $\nabla_\Phi Q'(\Phi|\Phi^{(n)}) < 0$ then λ is reduced towards zero; in this case the optimisation process operates in inverse Hessian mode.. When controlled effectively, this method is less prone to local convergence than the standard steepest gradient descent method, while offering efficiency gains over the inverse Hessian method.

3 Radar Data

The overall goal of the study reported in this paper is the registration of partial radar images against a digital map. Figure 1a shows a typical radar image. The

images are of rural areas in which the principal man-made linear-features available for matching are hedge-rows. The radar data is delivered as a series of non-overlapping sweeps interspersed with substantial dead-regions. Within each sweep there is both a significant oriented background texture and a systematic variation in background intensity. Before we can commence with the registration process, we must extract a set of line-segments for matching. Notwithstanding the difficulties associated with the partial sweeps, there are a number of obstacles to the characterisation of hedges from their radar reflections in speckle noise. In the first instance, the structures themselves are of variable width. Moreover, there are position and orientation dependant shadow artifacts which introduce an anti-symmetric component into the symmetric intensity profile of idealised hedge reflections. It is for these reasons that the hedge radar reflections are not well characterised by idealised even-symmetry fixed-width line-detectors such as the directional second derivative of Gaussian or the cosine-phase Gabor filter. Instead, we capture the variable width and mixed symmetry structure of the intensity profiles using a bank of sine and cosine phase Gabor filters of multiple orientation and scale. Here we adopt a statistical approach based on principal components analysis to find linear combination of these filter responses for optimal hedge-enhancement. Figure 1b shows the enhanced hedge-features obtained by applying the filter combinations to the raw radar image.

The next task is the re-enforcement of the filter responses to deliver connected linear-features. This second processing step is performed using a Bayesian relaxation scheme which utilises a statistical model of the filter-bank response [5]. The linear filter response combinations are used as input to a relaxation operator which iterates to enhance local connectivity using a dictionary of local line-structures [4]. Once stable line features are to hand, candidate line-segments may be extracted. Figure 1c shows the detected straight-line segments. Although most of the genuine hedge-features are detected, there is a significant population of extraneous lines.

4 Matching Experiments

We are now in a position to demonstrate the effectiveness of our registration process. Figure 2 shows the line registration process iterating from an initial guess to the final match. The red lines overlayed on the original radar image are the linear segments from the digital map. In this example, the initial parameters are misregistered in scale, orientation and position. The algorithm converges in approximately six iterations.

In iterating to the final solution, we have applied the Levenberg-Marquardt technique to the expected data log-likelihood $Q'(\Phi^{(n+1)}|\Phi^{(n)})$ to extract optimal parameter values at each iteration. One of the well-documented shortcomings of the EM algorithm is its local convergence properties. However, our motivation in adopting the Levenberg-Marquardt method in favour of straightforward steepest gradient have been to circimvent some of the problems of local convergence. In order to understand the limitations of our registration process, we have therefore investigated the structure of the log-likelihood function $Q'(\Phi^{(n+1)}|\Phi^{(n)})$ as a function of the scale, rotation and translation parameters. Figures 3a shows the expected log-likelihood as a function of scale and translation, while Figure 3b is a plot of expected log-likelihood as a function of rotation and scale. It is interesting

to note that while the scale-translation behaviour is unimodal, the rotation-scale plot contains local sub-optima. Closer inspection of the Figure 3b reveals that these sub-optima are associated with rotations of $\frac{\pi}{2}$. This rotation ambiguity is attributable to the fact that the hedge-rows in the digital map are organised into rectangular field-structures.

Our final set of experiments aims to evaluate the sensitivity of our method to random perturbations of the line-segments. Here we aim to illustrate the noise sensitivity of the match when the endpoints of the line-segments are subjected to Gaussian position errors of known variance. Figure 4 shows the fraction of lines correctly matched as a function of the line-end position error expressed as a fraction of the average interline spacing. The three curves correspond to different levels of initial line-displacement from the ground-truth solution. The upper curve corresponds to an average initial displacement which is 10% of the average inter-line spacing, in the case of the middle curve the displacement is 50% while in the case of the lower curve the displacement is 100%. The main conclusions from this plot is that the registration accuracy falls linearly with random measurement errors and that the method can not tolerate initialisation errors significantly greater then 50%.

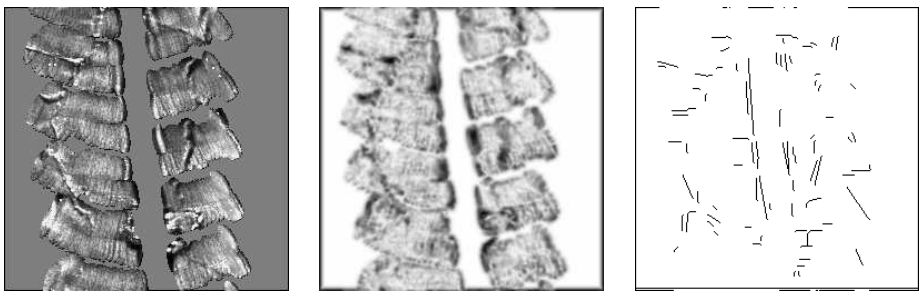

Figure 1: *Extracting line-segments from the radar data: a) Original radar image (note how the sweeps fragment the image features and the highly textured nature of image data); b) enhanced line-features after multi-channel filtering; c) the extracted line-segments*

5 Conclusions

We have detailed a technique for registering incomplete radar images which uses the EM algorithm to estimate transformation parameters using measurements provided by line-segments. The registration technique is only susceptible to local convergence if the initial rotation parameter is poorly estimated. In-fact, the data log- likelihood function is unimodal in scale and translation. We aim to minimise the difficulties associated with local convergence by adopting the Levenberg-Marquardt optimisation method. There are clearly a number of ways in which the work described in this paper can be extended. Although several authors have reported line registration algorithms [2, 11] these are invariably based on essentially *ad hoc* cost functions. One natural extension of our methodology is the recovery of 3D pose from 2D images [6]. Finally, the EM framework reported in this paper

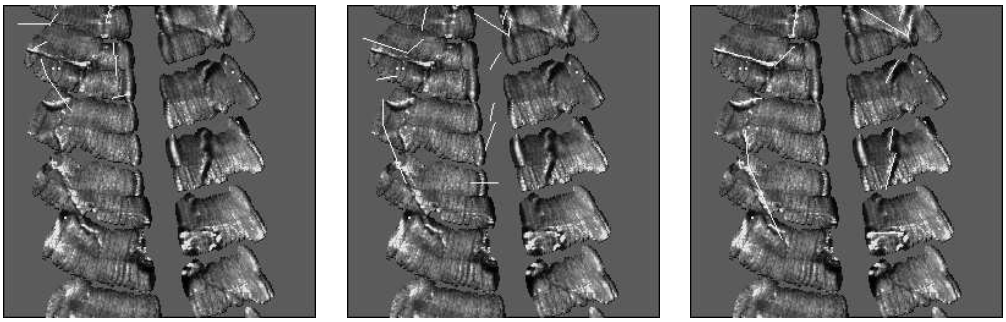

Figure 2: *Initial, third and final iterations of the map-fitting process. The algorithm converges after six iterations.*

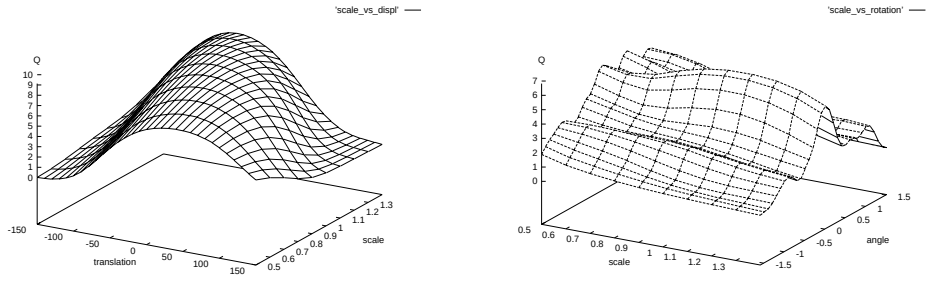

Figure 3: *The data log-likelihood function: a) as a function of translation and scale (note that the function is unimodal); b) as a function of scale and rotation (note that there are sub-optima associated with rotations of the model by angle multiples of* $\frac{\pi}{2}$*).*

is naturally extensible to multiple model matching [9].

References

[1] Becker S. and Hinton G.E, "Learning Mixture Models of Spatial Coherence", Neural Computation, **5**, pp 267–277, 1992.

[2] Beveridge J.R. and Riseman E.M., "Optimal Geometric Model-matching under Full 3D Perspective", *CVGIP: Image Understanding*, pp. 351–364, 1995.

[3] Dempster A.P. Laird N.M. and Rubin D.B., "Maximum-likelihood from Incomplete Data via the EM Algorithm", J. Royal Statistical Soc. Ser. B (methodological),**39**, pp 1-38, 1977.

[4] Evans A.N, Sharp N.G. and Hancock E.R., "Noise Models for Linear Feature Detection in SAR Images", *Proceedings of the 1994 IEEE International Conference on Image Processing*, **1**, pp. 466–470, 1994.

[5] Hancock E.R., "Resolving Edge-line Ambiguities using Probabilistic Relaxation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 300-306, 1993.

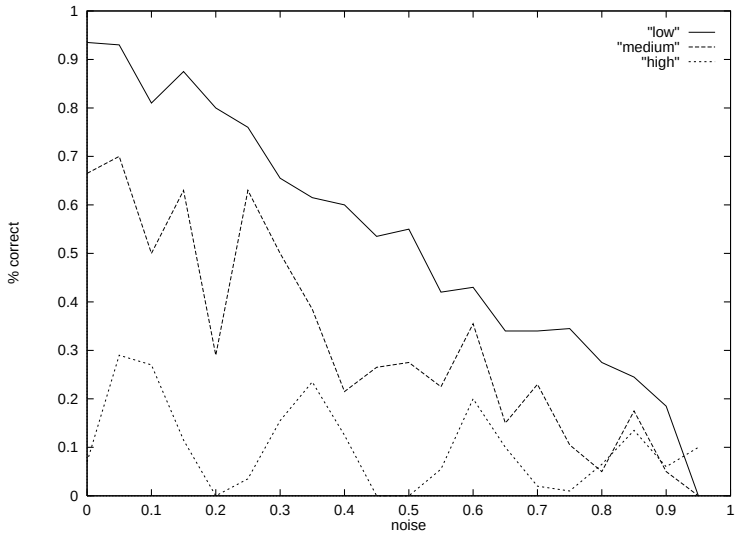

Figure 4: *Sensitivity to noise and displacement. From top to bottom, the three curves show the results of displacing the initial estimate by 10%, 50% and 100% of the average interline-spacing of the registered model. The noise is measured as the fractional random displacement error for the line-endings.*

[6] Hornegger J. and Niemann H., "Statistical Learning Localisation and Identification of Objects" *Proceedings Fifth International Conference on Computer Vision*, pp. 914–919, 1995.

[7] Jordan M.I. and Jacobs R.A, "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation*, **6**, pp. 181-214, 1994.

[8] Lavallee S. and Szeliski R., "Recovering the Position and Orientation of Free-form Objects from Image Contours using 3D Distance Maps", *IEEE PAMI*, **17**, pp. 378-390, 1995.

[9] Maclean J. and Jepson A,, "Recovery of Egomotion and Segmentation of Independent Object Motion using the EM Algorithm", *Proceedings BMVC*, pp. 175–184, 1994.

[10] Moghaddam B. and Pentland A., "Probabilistic Visual Learning for Object Detection", *Proceedings of the Fifth International Conference on Computer Vision*, pp. 786–793, 1995.

[11] Phong T.Q., Horaud R., Yassine A. and Tao P.D., "Object Pose from 2-D to 3-D Point and Line Correspondences", *International Journal of Computer Vision*, **15**, pp. 225–243, 1995.

[12] Wilson R.C., Evans A.N. and Hancock E.R., "Relational Matching by Discrete Relaxation", *Image and Vision Computing*, **13**, pp. 411-421, 1995.

[13] Viola P. and Wells W., "Alignment by Maximization of Mutual Information", *Proceedings of the Fifth International Conference on Computer Vision*, pp. 16–23, 1995.