

# Image Registration with Shape Mixtures

Simon Moss and Edwin R. Hancock

Department of Computer Science,  
University of York,  
York YO1 5DD, UK.  
erh@minster.york.ac.uk

**Abstract.** This paper describes how mixtures of Gaussians can be used for multiple shape template registration. The EM algorithm is applied to the shape mixture model to compute both maximum likelihood registration parameters together with a set of *a posteriori* matching probabilities. This architecture can be viewed as providing simultaneous registration and hypothesis verification. The different templates compete to account for data through the imposed probability normalisation. Based on a sensitivity study, our main conclusions are that the method is both robust to added image noise and poor initialisation.

## 1 Introduction

One of the key problems posed by large image data-bases is the need to recognise a particular object when multiple model hypotheses are to hand. Most vision systems accommodate the multiple hypothesis situation by invoking a two-step recognition loop. In the hypothesis generation step particular model poses are projected onto the raw image data. For reasons of efficiency, this step invariably draws on domain-specific cues to restrict the dimensionality of the search space (i.e. the number of model poses that must be projected onto the data) or limit the number of active hypotheses (i.e. to reject certain models as being unlikely to deliver good descriptions of the image data). In the verification step a measure of goodness of fit is used to assess how well model and data register against one-another. The best model is selected on the basis of this fitness measure.

Although this model-driven hypothesis and test approach can be rendered highly efficient, it clearly violates the principle of least commitment. Ambiguities are resolved early and partial matches are not fully evaluated. Moreover, there is no possibility for model overlap or mixing. Indeed, the absence of these possibilities is at odds with much of the psychology of model-recall [4,6]. In particular, it rules out important phenomenon such as perceptual alternation [1,11]. Our aim in this paper is to demonstrate that if efficient sequential search is not regarded as a computational imperative, then the connectionist literature can furnish more plausible algorithms for processing multiple model hypotheses. In particular aim to demonstrate that the mixture of Gaussians algorithm (MG) offers exciting possibilities for object-recognition from a data-base of multiple models.

The specific aim of this paper is to describe an application of the mixture of Gaussians algorithm [5] to the problem of registering an image against a large data-base of alternative models. The MG algorithm is a simplified version of the hierarchical mixture of experts algorithm of Jordan and Jacob [5] in which the supervisor layer is devoid of model-parameters. The algorithm can be regarded as an extension of the classical EM algorithm of Dempster, Laird and Rubin [3]. Basic to mixture of Gaussians is the idea of providing a mechanism for gating different competing model hypotheses. The associated *a posteriori* probabilities of model-match and the underlying model parameters are respectively computed using the expectation and maximisation steps of the classical EM algorithm. The expectation step of the algorithm can be viewed as combining evidence for the different competing model hypotheses in the data-base. The maximisation step is concerned with finding the maximum-likelihood parameters for each of the models. Because, each model has an associated probability there is scope for accommodating both ambiguity of appearance [2,10,6] and hypothesis alternation [1,11]. Moreover, different models can co-operate [7] to provide a partial explanation for the data.

We have recently exploited the EM algorithm to match individual cartographic models against very noisy milli-metre radar images [8,9]. The overall goal here has been to develop tool for aerial navigation. In this paper, we extend this earlier work to the case when a database of multiple templates is to hand. In other words, in addition to obtaining an accurate registration of the most appropriate model we also aim to assess the significance of a number of competing model hypotheses. The methodological basis for this study is the mixture of Gaussians algorithm [5]. Each shape template competes to explain the available image data. Associated with each model is a gating weight which is in reality the *a posteriori* model probability. Individual models also possess a set of maximum-likelihood registration parameters which describe the transformation between template co-ordinate system and the image co-ordinate system. Finally, there is an *a posteriori* probability of each of the data-items to have originated from the components of individual models. In our matching application these structural components are line-segments. Here the *a posteriori* probabilities represent the affinities between line segments in the data and those in the set of shape templates.

The outline of this paper is as follows. In Section 2 we furnish details of our line-based representation of model shapes. Section 3 briefly reviews the details of the EM algorithm that are pertinent to our study. Section 4 provides a sensitivity study aimed at demonstrating some of the properties of our matching method. Finally, Section 5 offers our conclusions and outlines our future plans.

## 2 Representation

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  $\theta_i$ . Each line in the image is

represented by a vector  $\mathbf{w}_i = (x_i, y_i, \theta_i)^T$  where  $i$  is the segment index. The complete set of data-lines is set  $\mathbf{w} = \{\mathbf{w}_i, \forall i \in \mathcal{D}\}$  where  $\mathcal{D}$  is the segment index-set. The data-base of models consists of a set of line-templates. Suppose that the template indexed  $\mu$  consists of the set of lines  $\mathbf{z}^\mu = \{\mathbf{z}_j^\mu, \forall j \in \mathcal{M}^\mu\}$ . Here  $\mathcal{M}^\mu$  is the index-set for the model lines and the  $\mathbf{z}_j^\mu$  represent the corresponding measurement-vectors. The aim of our matching algorithm is to iteratively recover a parameter-vector  $\phi_\mu^{(n)}$  which describes the Euclidean transformation that brings the data-lines into registration with the model indexed  $\mu$ . Each parameter-vector  $\phi_\mu^{(n)} = (\phi_{\mu,1}^{(n)}, \dots, \phi_{\mu,4}^{(n)})^T$  has four components; the meanings of these components are as follows;  $\phi_{\mu,1}^{(n)}$  represents the x-translation,  $\phi_{\mu,2}^{(n)}$  represents the y-translation,  $\phi_{\mu,3}^{(n)}$  is the rotation and  $\phi_{\mu,4}^{(n)}$  is the relative scale. The state of registration of the entire model-base is denoted by the matrix of column vectors  $\Phi^{(n)} = (\phi_1^{(n)}, \dots, \phi_K^{(n)})$  where  $K$  is the number of models. Individual model-lines are transformed into the co-ordinate system of the image by applying the appropriate set of Euclidean parameters. The transformed position and orientation for the  $j$ th line in the  $\mu$ th model is given by

$$F(\mathbf{z}_j^\mu, \phi_\mu^{(n)}) = U(\phi_\mu^{(n)})\mathbf{z}_j^\mu + V\phi_\mu^{(n)} \quad (1)$$

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

$$U(\phi_\mu^{(n)}) = \begin{pmatrix} \phi_{\mu,4}^{(n)} \cos \phi_{\mu,3}^{(n)} & -\phi_{\mu,4}^{(n)} \sin \phi_{\mu,3}^{(n)} & 0 \\ \phi_{\mu,4}^{(n)} \sin \phi_{\mu,3}^{(n)} & \phi_{\mu,4}^{(n)} \cos \phi_{\mu,3}^{(n)} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

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

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (3)$$

### 3 The Mixture of Gaussians 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 of Dempster, Laird and Rubin [3]. The algorithm recovers multiple model descriptions in a two stage process. 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. In our application the different models are sets of line-segments residing within a data-base of possible image interpretations.

Basic to our philosophy of exploiting the EM algorithm is the idea that every line-segment in the the set of available templates can in principle associate to each of the lines in each of stored models 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 ingredients of the mixture density are the component conditional measurement densities  $p(\mathbf{w}_i | \mathbf{z}_j^\mu, \phi_\mu^{(n)})$ , together with the mixing proportions of specific model lines and the different model mixing proportions. The conditional measurement densities represent the likelihood that the data-line measurement  $\mathbf{w}_i$  originates from the line indexed  $j$  for the model indexed  $\mu$  under the prevailing set of transformation parameters  $\phi_\mu^{(n)}$  for that particular model. We use the shorthand notation  $\alpha_{j,\mu}^{(n)}$  to represent the mixing proportion for the line  $j$  from the model  $\mu$ . The overall proportion of the model  $\mu$  at iteration  $n$  is  $\beta_\mu^{(n)}$ . These quantities provide a natural mechanism for assessing the significance of the individual lines and models in explaining the current data-likelihood. For instance if  $\alpha_{j,\mu}^{(n)}$  approaches zero, then this indicates that there is no line in the data that matches the line  $j$  in the model  $\mu$ . In other words, the mixture model provides a natural way not only of accommodating missing segments but also for handling model overlap. It is important to stress that the mixing proportions are iteration dependent, being conditioned upon the current parameter values.

### 3.1 Expectation

The expectation step of the EM algorithm is aimed at estimating the data 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 in the image and their counterparts in the set of stored models. 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, is equivalent to maximising the weighted log-likelihood function

$$Q(\Phi^{(n+1)} | \Phi^{(n)}) = \sum_{\mu=1}^K \sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{M}^\mu} P(\mathbf{z}_j^\mu | \mathbf{w}_i, \phi_\mu^{(n)}) \ln p(\mathbf{w}_i | \mathbf{z}_j^\mu, \phi_\mu^{(n+1)}) \quad (4)$$

The *a posteriori* probabilities  $P(\mathbf{z}_j^\mu | \mathbf{w}_i, \phi_\mu^{(n)})$  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  $j$  from the model indexed  $\mu$ . In other words, they represent model-datum affinities. Using the Bayes rule, we re-write the *a posteriori* matching probabilities in terms of the conditional measurement densities

$$P(\mathbf{z}_j^\mu | \mathbf{w}_i, \phi_\mu^{(n)}) = \frac{\beta_\mu^{(n)} \alpha_{j,\mu}^{(n)} p(\mathbf{w}_i | \mathbf{z}_j^\mu, \phi_\mu^{(n)})}{\sum_{\mu'} \sum_{j' \in \mathcal{M}^{\mu'}} \beta_{\mu'}^{(n)} \alpha_{j',\mu'}^{(n)} p(\mathbf{w}_i | \mathbf{z}_{j'}^{\mu'}, \phi_{\mu'}^{(n)})} \quad (5)$$

The line mixing proportions for each model in turn are computed by averaging the *a posteriori* probabilities over the set of data-lines, i.e.

$$\alpha_{j,\mu}^{(n+1)} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} P(\mathbf{z}_j^\mu | \mathbf{w}_i, \phi_\mu^{(n)}) \quad (6)$$

The *a posteriori* model probabilities are found by summing the relevant set of line mixing proportions, i.e.

$$\beta_\mu^{(n+1)} = \sum_{j \in \mathcal{M}^\mu} \alpha_{j,\mu}^{(n+1)} \quad (7)$$

In this way the *a posteriori* model probabilities sum to unity over the complete set of models. The probability assignment scheme allows for both model overlap and the assessment of ambiguous hypotheses.

To proceed we require a model for the conditional measurement densities, i.e.  $p(\mathbf{w}_i | \mathbf{z}_j^\mu, \phi_\mu^{(n)})$ . 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  $j$ th model line delivered by the current estimated transformation parameters. Accordingly we write

$$p(\mathbf{w}_i | \mathbf{z}_j^\mu, \phi_\mu^{(n)}) = \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{|\Sigma_\mu|}} \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n)})^T \Sigma_\mu^{-1} \boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n)}) \right] \quad (8)$$

In the above expression  $\Sigma_\mu$  is the variance-covariance matrix for the model indexed  $\mu$ . The quantity  $\boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n)}) = \mathbf{w}_i - F(\mathbf{z}_j^\mu, \phi_\mu^{(n)})$  is the vector of error-residuals between the data-line indexed  $i$  and transformation of the line indexed  $j$  in the model indexed  $\mu$  under the current set of Euclidean parameters, i.e.  $\phi_\mu^{(n)}$ . The variance-covariance matrix for the model indexed  $j$  is estimated from the expected value of the error-residuals, i.e.,  $\Sigma_\mu = E[\boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n)}) \cdot \boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n)})^T]$ . Initially, we commence with a diagonal estimate of the co-variance matrix with  $x$  and  $y$  components set equal to one-another.

### 3.2 Maximisation

With the ingredients outlined in the previous subsection, the expectation step of the EM algorithm simply reduces to computing the weighted squared error criterion

$$Q'(\Phi^{(n+1)} | \Phi^{(n)}) = -\frac{1}{2} \sum_{\mu=1}^K \sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{M}} P(\mathbf{z}_j^\mu | \mathbf{w}_i, \phi_\mu^{(n)}) \boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n+1)})^T \Sigma_\mu^{-1} \boldsymbol{\epsilon}_{i,j,\mu}(\phi_\mu^{(n+1)}) \quad (9)$$

In other words, the *a posteriori* probabilities  $\beta_\mu^{(n)}$  and  $\alpha_{j,\mu}^{(n)}$  effectively regulate the contributions to the likelihood function. Models or lines for which there is little evidence of match contribute insignificantly, while those which are in

good registration dominate. The maximisation step aims to locate the updated matrix of parameter-vectors  $\Phi^{(n+1)}$  that optimises the quantity  $Q'(\Phi^{(n+1)}|\Phi^{(n)})$ , i.e.

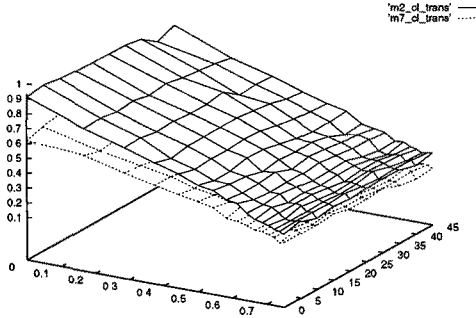
$$\Phi^{(n+1)} = \arg \max_{\Phi} Q'(\Phi|\Phi^{(n)}) \quad (10)$$

We solve the implied weighted least-squares minimisation problem using the Levenberg-Marquardt technique. 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. The main advantages are speed of convergence and a reduced susceptibility to local optima. Full details of the optimisation step can be found in the recent account of Moss and Hancock [9].

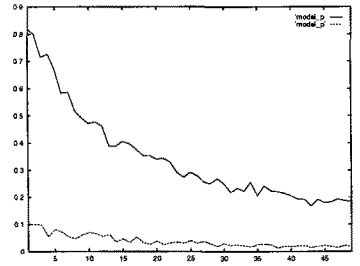
## 4 Experiments

The overall goal of the study reported in this paper is to demonstrate that the mixture of Gaussians algorithm can be used to simultaneously recognize and register models from a data-base. The recognition process is provided by the expectation step of the algorithm in which model gating probabilities are updated. Registration takes place in the maximisation step where maximum likelihood transformation parameters are estimated from the available image data. We have performed a sensitivity analysis with the aim of showing the effect of noise and clutter on the model data matching probabilities, and to show what effect the number of models stored has on these probabilities. In order to do this, we have constructed a number of synthetic problems composed of randomly placed and orientated line segments. Each model set, contains only one model which can be accurately registered to the data set. Our first experiment, consists of varying the amount of noise on the data segments and the addition of clutter line segments. Figure 1a shows the 2-d surface of matching probabilities as a function of these two quantities. In this case, noise takes the form of random displacement of the data line segments measured in terms of the standard deviation of the Gaussian noise. It can be seen that the addition of clutter causes the matching probability to steadily decrease. This can be attributed to the clutter segments forming spurious local minima in the energy function, corresponding to false matches. Clearly the larger the number of clutter segments, the larger the likelihood of false correspondences and hence the lower the probability of a model-data assignment. The relative insensitivity of the method to translational noise can be attributed to the inclusion of line orientation in the cost function and it's weighting through the use of the covariance matrix  $\Sigma$ .

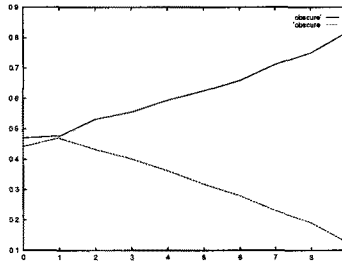
Figure 1b shows how the model matching probability varies as a function of the number of models stored by the system. The green line corresponds to the probability of the true model matching the data, whilst the lower line shows the highest matching probability from any of the remaining models. It can be seen that the true probability falls rapidly as the number of models increases.



(a) Effect of noise and clutter.



(b) Effect of number of models.



(c) The effect of model overlap.

**Fig. 1.** Sensitivity plots for the gating probabilities.

However, even though the matching probability decreases rapidly, it remains higher than any of the model matching probabilities. In this case, taking the model with the *maximum a posteriori* matching probability always results in the correct model classification. However, it should be noted that this does not guarantee that the final transformation parameters are accurate.

Finally, we provide some experiments on the effect of model overlap. We have generated a scene which consists of two noise corrupted models from our database. We have subsequently systematically deleted a fraction of the lines from one of the models. Figure 1c shows the gating probabilities for the two models as a function of the fraction of deleted line-segments. As the fraction increases so the gating probability of the in-tact model increases while the gating probability of the corrupted model decreases. Moreover the gating probabilities directly reflect the mixing fractions of the two sets of lines.

## 5 Conclusions

Our main contribution in this paper has been to show how the mixture of Gaussians algorithm can be used as a tool for simultaneous model recognition and registration. Our contribution can be regarded as illustrating the effectiveness of the EM algorithm in recalling models stored in a data-base and registering them against noisy image data. In other words, the algorithm has significant potential for shape-recall from image data-bases. There are clearly ways in which the current work can be extended. In the first instance, it would be interesting to know the shape memory capacity of our algorithm. Moreover, there is still considerable scope for assessing the capacity of the technique to handle scenes which contain multiple instances of overlapped models. Studies aimed at answering these questions are in progress and will be reported in due course.

## References

1. Bialek W. and M. Deweese, "Random switching and optimal processing in the perception of ambiguous figures- a neural network model", *Physical Review Letters*, **74**, pp. 3077-3080, 1995.
2. Callari F.G. and F.P. Ferrie, "Active recognition: using uncertainty to reduce ambiguity", *Proceedings IEEE CVPR Conference*, pp. 701-707, 1996
3. Dempster A.P., Rubin 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. Feldman J.A. and D.H. Ballard, "Connectionist models and their Properties", *Cognitive Science*, **6**, pp. 205-254, 1982.
5. Jordan M.I. and Jacobs R.A., "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation*, **6**, pp. 181-214, 1994.
6. Kawabata N. and T. Mori, "Disambiguating ambiguous figures by a model of selective attention", *Biological Cybernetics*, **67**, pp. 417-425, 1992.
7. Minka T. and Picard R., "Interactive Learning with a Society of Models", *IEEE Computer Society Computer Vision and Pattern Recognition Conference*, pp. 447-452, 1996.
8. Moss S. and E.R. Hancock, "Registering Incomplete Radar Images using the EM Algorithm", *Proceedings of the British Machine Vision Conference*, pp. 685-694, 1996.
9. Moss S. and E.R. Hancock, "Cartographic Matching onto Millimetre Radar Images", *IEEE Computer Society Workshop on Applications of Computer Vision*, 1996.
10. Qin C. and J.Y.S. Luh, "Ambiguity Reduction in Relaxation Labelling", *Pattern Recognition*, **27**, pp. 165-180, 1994.
11. Riani M. and E. Simonotto, "Stochastic resonance in the perceptual interpretation of ambiguous figures - a neural network approach", *Physical Review Letters*, **72**, pp. 3120-3123, 1994.
12. Yuille A.L., Stolorz P. and Utans J., "Statistical Physics, Mixtures of Distributions, and the EM Algorithm" *Neural Computation*, **6**, pp. 334-340, 1994.