Holistic Matching

Andrew D. J. Cross and Edwin R. Hancock

Department of Computer Science
University of York
York, Y01 5DD, UK
email: erh,adjc@minster.york.ac.uk

Abstract. This paper describes a new approach to extracting affine structure from 2D point-sets. The novel feature is to unify the tasks of estimating transformation geometry and identifying point-correspondence matches. Unification is realised by constructing a mixture model over the bi-partite graph representing the correspondence match and by effecting optimisation using the EM algorithm. According to our EM framework the probabilities of structural correspondence gate contributions to the expected likelihood function used to estimate maximum likelihood affine parameters. This provides a means of rejecting structural outliers. We evaluate the technique on the matching of different affine views of 3.5in floppy discs. We provide a sensitivity study based on synthetic data.

1 Introduction

The estimation of transformational geometry is key to many problems of computer vision and robotics [11, 12]. Broadly speaking the aim is to recover a matrix representation of the transformation between image and world co-ordinate systems. In order to estimate the matrix requires a set of correspondence matches between features in the two co-ordinate systems [13]. Posed in this way there is a basic chicken-and-egg problem. Before good correspondences can be estimated, there need to be reasonable bounds on the transformational geometry. Yet this geometry is, after all, the ultimate goal of computation. This problem is usually overcome by invoking constraints to bootstrap the estimation of feasible correspondence matches [6, 9]. One of the most popular ideas is to use the epipolar constraint to prune the space of potential correspondences [6].

The aim in this paper is to develop a synergistic or holistic framework for matching. Specifically, we aim to facilitate feedback between the two problems of estimating transformational geometry and locating correspondence matches. We realise this goal using an architecture that is reminiscent of the hierarchical mixture of experts algorithm [7]. The key idea is to use a bi-partite graph to represent the current configuration of correspondence match. This graphical structure provides an architecture that can be used to gate contributions to the likelihood function for the geometric parameters using structural constraints. Correspondence matches and transformation parameters are estimated by applying the EM algorithm to the gated likelihood function. In this way we arrive

at dual maximisation steps. Maximum likelihood parameters are found by minimising the structurally gated squared error residuals between features in the two images being matched. Correspondence matches are updated so as to maximise the *a posteriori* probability of the observed structural configuration on the bi-partite association graph.

It is important to stress that the idea of using a graphical model to provide structural constraints on parameter estimation is a task of generic importance. Although the EM algorithm has been used to extract affine and Euclidean parameters from point-sets [15, 5] or line-sets [10], there has been no attempt to impose structural constraints of the correspondence matches. Viewed from the perspective of graphical template matching [1, 8] our EM algorithm allows an explicit deformational model to be imposed on a set of feature points. Since the method delivers statistical estimates for both the transformation parameters and their associated covariance matrix it offers significant advantages in terms of its adaptive capabilities.

We provide a practical illustration for several matching tasks. These examples include the matching of images of floppy discs. Here we illustrate that the technique degrades gracefully even when there is severe perspective foreshortening. The second example focusses on aerial images where there is significant barrel distortion.

2 Affine Transformation of Point Sets

Our goal is to recover the parameters of a geometric transformation $\Phi^{(n)}$ that best maps a set of image feature points $\mathbf{w}$ onto their counterparts in a model $\mathbf{z}$. In order to do this, we represent each point in the image data set by an augmented position vector $\boldsymbol{w}_i = (x_i, y_i, 1)^T$ where i is the point index. This augmented vector represents the two-dimensional point position in a homogeneous coordinate system. We will assume that all these points lie on a single plane in the image. In the interests of brevity we will denote the entire set of image points by $\mathbf{w} = \{\boldsymbol{w}_i, \forall i \in \mathcal{D}\}$ where $\mathcal{D}$ is the point set. The corresponding fiducial points constituting the model are similarly represented by $\mathbf{z} = \{\boldsymbol{z}_j, \forall j \in \mathcal{M}\}$ where $\mathcal{M}$ denotes the index-set for the model feature-points $\boldsymbol{z}_j$.

In this paper we are interested in affine transformations. The affine transformation has six free parameters. These model the two components of translation of the origin on the image plane, the overall rotation of the co-ordinate system, the overall scale together with the two parameters of shear. These parameters can be combined succinctly into an augmented matrix that takes the form

$$\Phi^{(n)} = \begin{pmatrix} \phi_{1,1}^{(n)} & \phi_{1,2}^{(n)} & \phi_{1,3}^{(n)} \\ \phi_{2,1}^{(n)} & \phi_{2,2}^{(n)} & \phi_{2,3}^{(n)} \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

With this representation, the affine tranformation of co-ordinates is computed using the following matrix multiplication

$$z_j^{(n)} = \Phi^{(n)} z_j \tag{2}$$

Clearly, the result of this multiplication gives us a vector of the form $z_j^{(n)} = (x, y, 1)^T$. The superscript n indicates that the parameters are taken from the n^{th} iteration of our algorithm. Our goal is to recover the elements $\phi_{i,j}^{(n)}$ of the parameter matrix $\Phi^{(n)}$, which describes a coordinate system transformation that will best bring the set image points $\mathbf{w}$ into registration with the model set $\mathbf{z}$.

Since the affine transformation can be represented in a linear fashion, the parameter recovery process is easily realised by matrix inversion. To be fully constrained, the recovery process requires three image points that are known to be in correspondence. There are several ways in which the correspondences may be established. One of the most popular is to use the epipolar constraint to search for candidate matches prior to parameter recovery. If more than three such points are available, then the parameter recovery process is over-constrained, and can be realised by least-squares fitting. If reliable correspondences are not available, then a robust fitting method must be employed. This involves removing rogue correspondences through outlier reject. A concrete example is furnished by the recent work of Torr [13].

In this paper we adopt a somewhat different approach to the affine recovery problem. We take the view that the available correspondences are at best uncertain and may contain a substantial proportion of errors. However, rather than rejecting those correspondences which give rise to a large affine registration error, we attempt to iteratively correct them. In a nutshell, our idea is to alternate between estimating affine parameters and refining correspondence matches. The framework for this study is furnished by a variant of the EM algorithm. Specifically, we use a gating process similar to that of Jordan and Jacob's [7] hierarchical mixture of experts architecture to control contributions to the log-likelihood function for the affine parameters.

The gating layer represents the state of correspondence match between the point-sets. Rather than using epipolar constraints to gauge consistency, we use constraints provided by the spatial proximity of the points. These constraints are elicited by separately triangulating the data and model points. The proximity constraints provided by the correspondences between two edges weight the contributions to the log-likelihood function. In the next Section we describe how the relational consistency of the correspondence match can be modelled in a probabilistic manner.

3 Relational Graph Matching

One of our goals in this paper is to exploit structural constraints to improve the recovery of transformation parameters from sets of feature points. We abstract the process as bi-partite graph matching. Because of its well documented robustness to noise and change of viewpoint, we adopt the Delaunay triangulation as

our basic representation of image structure [14, 4]. We establish Delaunay triangulations on the data and the model, by seeding Voronoi tessellations from the feature-points.

The process of Delaunay triangulation generates relational graphs from the two sets of point-features. More formally, the point-sets are the nodes of a data graph $G_D = \{\mathcal{D}, E_D\}$ and a model graph $G_M = \{\mathcal{M}, E_M\}$. Here $E_D \subseteq \mathcal{D} \times \mathcal{D}$ and $E_M \subseteq \mathcal{M} \times \mathcal{M}$ are the edge-sets of the data and model graphs. Key to our matching process is the idea of using the edge-structure of Delaunay graphs to constrain the correspondence matches between the two point-sets. This correspondence matching is denoted by the function $f : \mathcal{M} \rightarrow \mathcal{D}$ from the nodes of the data-graph to those of the model graph. According to this notation the statement $f^{(n)}(i) = j$ indicates that there is a match between the node $i \in \mathcal{D}$ of the model-graph to the node $j \in \mathcal{M}$ of the data-graph at iteration n of the algorithm. We use the binary indicator

$$s_{i,j}^{(n)} = \begin{cases} 1 & \text{if } f^{(n)}(i) = j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

to represent the configuration of correspondence matches.

3.1 Relational Consistency

We exploit the structure of the Delaunay graphs to compute the consistency of match using the Bayesian framework for relational graph-matching recently reported by Wilson and Hancock [16]. Details are beyond the scope of this paper. Suffice to say that consistency of a configuration of matches residing on the neighbourhood $R_i = i \cup \{k \ ; \ (i,k) \in E_D\}$ of the node i in the data-graph and its counterpart $S_j = j \cup \{l \ ; \ (j,l) \in E_m\}$ for the node j in the model-graph is gauged by Hamming distance. The Hamming distance $H(i,j)$ counts the number of matches on the data-graph neighbourhood R_i that are inconsistently matched onto the model-graph neighbourhood S_j. According to Wilson and Hancock [16] the structural probability for the correspondence match $f(i) = j$ at iteration n of the algorithm is given by

$$\zeta_{i,j}^{(n)} = \frac{\exp\Big[-\beta H(i,j)\Big]}{\sum_{j \in \mathcal{M}} \exp\Big[-\beta H(i,j)\Big]} \tag{4}$$

In the above expression, the Hamming distance is given by

$$H(i,j) = \sum_{(k,l) \in R_i \bullet S_j} (1 - s_{k,l}^{(n)}) \tag{5}$$

where the symbol $\bullet$ denotes the composition of the data-graph relation R_i and the model-graph relation S_j. The exponential constant $\beta = \ln \frac{1-P_e}{P_e}$ is related to the uniform probability of structural matching errors P_e. This probability is set to reflect the overlap of the two point-sets. In the work reported here we set $P_e = \frac{2||\mathcal{M}|-|\mathcal{D}||}{||\mathcal{M}|+|\mathcal{D}||}$.

4 The Holistic Matching Algorithm

Our aim is to extract affine pose parameters and correspondences matches from the two point-sets using the EM algorithm. When couched probabilistically, the goal can be succinctly stated as that of jointly maximising the data-likelihood $p(\mathbf{w}|\mathbf{z}, f, \Phi)$ over the space of correspondence matches f and the matrix of affine parameters Φ. We realise this process using a dual-step or hierarchical version of the EM algorithm. According to the original work of Dempster *et al* [3] the expected likelihood function is computed by weighting the current log-probability density by the *a posteriori* measurement probabilities computed from the preceding maximum likelihood parameters. Here we wish to exploit Jordan and Jacobs [7] idea of augmenting the maximum likelihood process with a graphical model. From an architectural standpoint, the graphical model can be regarded as a supervisor network which effectively gates contributions to the expected log-likelihood function. The novelty of the work reported here is to develop a variant of this idea in which it is the bi-partite graph, i.e. f, which gates the likelihood function for the affine parameters Φ. This graph represent the current state of correspondence match between the two point-sets.

We extract both maximum likelihood perspective parameters and maximum *a posteriori* matching probabilities by applying coupled update operations to the gated likelihood function. In this way the consistency of the structural matching process can guide the pose recovery process. Likewise error probabilities derived from the position residuals are used to guide the correspondence matching process. When the joint likelihood function is maximised in this way, when the correspondence matches play the role of missing data.

In the spirit of Dempster, Laird and Rubin's EM algorithm [3], we aim to condition the updated parameter estimates (i.e. $\Phi^{(n+1)}$) on the most recently available correspondence matches (i.e. $f^{(n)}$). In other words, the maximum-likelihood parameters satisfy the following condition

$$\Phi^{(n+1)} = \arg\max_{\Phi} p(\Phi|\mathbf{w}, f^{(n)}) \tag{6}$$

In a similar way, the *maximum a posteriori* matches are conditioned upon the most recently available parameter-estimates. The matching configuration therefore satisfies the following condition

$$f^{(n+1)} = \arg\max_{f} P(f|\mathbf{w}, \Phi^{(n)}) \tag{7}$$

4.1 The gated likelihood function

We have recently shown how coupled updates of this form can be realised through the optimisation of single integrated expected likelihood function. Details of the formal development are outside the scope of this paper and can found in the recent account of Cross and Hancock [2]. Suffice to say that the parameters and the correspondence matches may be sought through joint optimisation of the quantity

$$Q(\Phi^{(n+1)}|\Phi^{(n)}) = \sum_{(i,j)\in f^{(n)}} P(\boldsymbol{z}_j|\boldsymbol{w}_i,\Phi^{(n)})\zeta_{i,j}^{(n)} \ln p(\boldsymbol{w}_i|\boldsymbol{z}_j,\Phi^{(n+1)}) \quad (8)$$

The structure of this expected log-likelihood function requires further comment. The measurement densities $p(\boldsymbol{w}_i|\boldsymbol{z}_j,\Phi^{(n+1)})$ model the distribution of error-residuals between the observed model-point position $\boldsymbol{w}_i$ and the predicted position of the model point $\boldsymbol{z}_j$ under the current set of transformation parameter $\Phi^{(n+1)}$. The log-likelihood contributions at iteration $n+1$ are weighted by the *a posteriori* measurement probabilities $P(\boldsymbol{z}_j|\boldsymbol{w}_i,\Phi^{(n)})$ computed at the previous iteration n of the algorithm. Following Jordan and Jacobs [7] we gate the individual expected-likelihood contributions using the the structural matching probabilities $\zeta_{i,j}^{(n)}$. Finally, the summation extends over the set of correspondence matches $(i,j) \in f^{(n)}$ available at iteration n.

Using the Bayes rule, we can re-write the *a posteriori* measurement probabilities in terms of the of the conditional measurement densities

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

The mixing proportions $\alpha_j^{(n)}$ are computed by averaging the *a posteriori* probabilities over the set of data-points, i.e.

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

In order to proceed with the development of a point registration process we require a model for the conditional measurement densities, i.e. $p(\boldsymbol{w}_i|\boldsymbol{z}_j,\Phi^{(n)})$. 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 predictions of the jth model point delivered by the current estimated transformation parameters. Accordingly we write

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

In the above expression Σ is the variance-covariance matrix for the vector of error-residuals $\epsilon_{i,j}(\Phi^{(n)}) = \boldsymbol{w}_i - \boldsymbol{z}_j^{(n)}$ between the components of the predicted measurement vectors $\boldsymbol{z}_j^{(n)}$ and their counterparts in the data, i.e. $\boldsymbol{w}_i$. Formally, the matrix is related to the expectation of the outer-product of the error-residuals i.e. $\Sigma = E[\epsilon_{i,j}(\Phi^{(n)})\epsilon_{i,j}(\Phi^{(n)})^T]$. With these ingredients, the expectation step of the EM algorithm simply reduces to computing the weighted squared error criterion

$$Q'(\Phi^{(n+1)}|\Phi^{(n)}) = \sum_{(i,j)\in f^{(n)}} P(\boldsymbol{z}_j|\boldsymbol{w}_i,\Phi^{(n)})\zeta_{i,j}^{(n)}\epsilon_{i,j}(\Phi^{(n+1)})^T\Sigma^{-1}\epsilon_{i,j}(\Phi^{(n+1)}) \quad (12)$$

In other words, the *a posteriori* probabilities $P(\boldsymbol{z}_j|\boldsymbol{w}_i, \Phi^{(n)})$ and the structural matching probabilities $\zeta_{i,j}^{(n)}$ 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.

4.2 Maximisation

As pointed out earlier, the maximisation step of our holistic matching algorithm is based on dual coupled update processes. The first of these aims to locate maximum *a posteriori* probability correspondence matches. The second update operation is concerned with locating maximum likelihood perspective parameters. We effect the coupling by allowing information flow between the two processes.

Maximum *a posteriori* probability matches: Point correspondences are sought so as to maximise the *a posteriori* probability of structural match. Individual point-correspondences should be updated in the following manner

$$f^{(n+1)}(i) = \arg\max_{j \in \mathcal{M}} P(\boldsymbol{z}_j|\boldsymbol{w}_i, \Phi^{(n)})\zeta_{i,j}^{(n)} \tag{13}$$

Once this update equation has been applied, the unmatched model-graph nodes are identified for removal from the triangulation. At this point the edited set of model feature-points is re-triangulated to correct potential structural errors. We provide more detials of this graph-editing process in Section 4.2.3. A full account of the method can be found in the recent paper of Wilson and Hancock [16]. The updated structural matching probabilities $\zeta_{i,j}^{(n+1)}$ are also updated using equations (4) and (5) as outlined in Section 3.

Maximum likelihood parameters: In the case of affine geometry, the transformation is linear in the parameters. This allows us to locate the maximum-likelihood parameters directly by solving the following system of saddle-point equations for the independent affine parameters $\phi_{k,l}^{(n+1)}$ running over the indices $k = 1, 2$ and $l = 1, 2, 3$

$$\frac{\partial Q'(\Phi^{(n+1)}|\Phi^{(n)})}{\partial \phi_{k,l}^{(n+1)}} = 0 \tag{14}$$

For the affine transformation the set of saddle-point equations is linear, and are hence easily solved by using matrix inversion. It is a straightforward to show that the updated matrix of affine parameters must satisfy the following implied system of linear equations

$$\sum_{(i,j) \in f^{(n)}} P(\boldsymbol{z}_j|\boldsymbol{w}_i, \Phi^{(n)})\zeta_{i,j}^{(n)} \left[(\boldsymbol{w}_i - \Phi^{(n+1)}\ \boldsymbol{z}_j)^T \Sigma^{-1} \right] \boldsymbol{z}_j\ U = 0 \tag{15}$$

where the elements of the matrix U are the partial derivatives of the affine transformation matrix with respect to the individual parameters, i.e.

$$U = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{16}$$

As a result the updated solution matrix is given by

$$\begin{aligned} \Phi^{(n+1)} = & \left[\sum_{(i,j) \in f^{(n)}} P(\boldsymbol{z}_j | \boldsymbol{w}_i, \Phi^{(n)}) \zeta_{i,j}^{(n)} \boldsymbol{z}_j \; U^T \; \boldsymbol{z}_j^T \; \Sigma^{-1} \right]^{-1} \\ & \times \left[\sum_{(i,j) \in f^{(n)}} P(\boldsymbol{z}_j | \boldsymbol{w}_i, \Phi^{(n)}) \zeta_{i,j}^{(n)} \boldsymbol{w}_i \; U^T \; \boldsymbol{z}_j^T \; \Sigma^{-1} \right] \end{aligned} \tag{17}$$

This allows us to recover a set of improved transformation parameters at iteration $n+1$. Once these are computed, the *a posteriori* measurement probabilities may be updated by applying the Bayes formula to the measurement density function. The update procedure involves substituting the parameter matrix of equation (1) into the Gaussian density of equation (11) and applying the Bayes theorem.

Updating the triangulation: Once we have an estimate of the transformation parameters for iteration n, we can use these to project the model point set $\mathbf{z}$ onto the data point set $\mathbf{w}$ using the recovered transformational geometry as outlined in section 2. Our Delaunay graph structures are completely invariant to translations, uniform scalings and rotations. They are in addition robust to the effects of non-uniform scaling, shear and perspective foreshortening. However under severe deformations their structures do become perturbed. In order to overcome this source of potential structural corruption, at the end of each iteration we re-triangulate the graph in order to accurately reflect the structure of the points under the current estimate of the transformation parameters.

In addition to this structural modification, we can improve the robustness of parameter estimation by removing points in the model-set which have no correspondence in the data-set when computing the expected log-likelihood function in the expectation step of the EM algorithm. Once these points are removed we must once again re-triangulate the point set in order to reflect the change in structure. At each iteration of the maximisation stage, we also try re-introducing any deleted points back into the data set.

5 Results

In this section, we will provide experimental evaluation of our new coupled matching process. This investigation has two distinct directions. Firstly, we will experimentally compare our algorithm with some commonly used alternatives.

In particular we will compare our scheme with standard least squares parameter recovery. We will also make comparison with a purely structural correspondence matching scheme. This first pair of comparisons serve to demonstrate that the combined modeling of both point correspondences and transformation geometry yields significant advantages in terms of accuracy of convergence over their individual use. In other words, there are tangible advantages to our holistic approach. In the second part of the experimental investigation we will furnish examples showing the use of the holistic scheme on real world imagery. Here we will use two different data sets. The first of these involves perspective views of 3.5 inch floppy discs. The second example involves matching distorted aerial image data against a digital map.

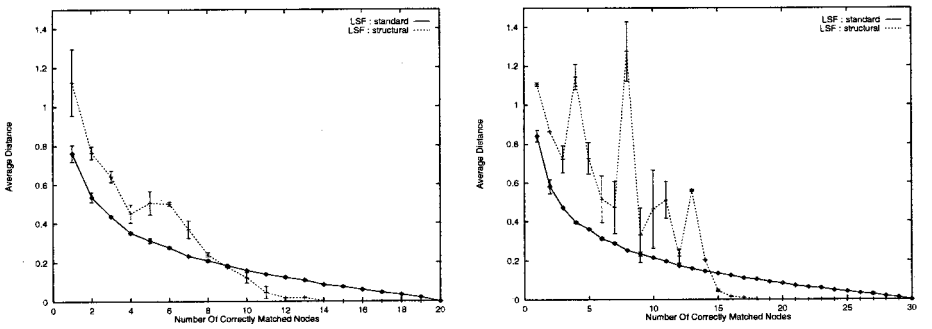
5.1 Algorithm Comparison

The aim of this section is to demonstrate how the holistic matching algorithm performs in comparison with other, similar, schemes reported in the literature. The principal modes of comparison will be with

- **Standard Least Squares Fitting:** The aim of this investigation is to demonstrate that the structural component of the model can make a significant impact on the robust recovery of affine parameters. In this section we will demonstrate how, by making no other modifications to the standard expectation-maximisation scheme, other than weighting the contributions of the different feature-points according by their structural consistency, we can recover the parameters in a robust manner. This performance advantage persists right down to the limits imposed by the number of degrees of freedom of the projective transformation.
- **Standard Structural Matching:** In this set of experiments we aim to demonstrate how our EM methodology performs in comparison with standard structural matching [16]. Viewed from an alternative perspective, our EM approach can equivalently be seen as a natural way of weighting standard structural matching schemes using a model of the point set transformation. When viewed in this way, it is clearly important to investigate the role played by the explicit modeling of the transformational geometry.

The results of the comparative study have been obtained using random point sets. This allows us to compare algorithm sensitivity in a controlled manner under varying noise conditions. This experimental methodology also allows the results to be averaged over a large number of runs, and meaningful error bars to be derived from the whole population of trials. In each of the following experiments, 100 trials where made for each point on the graph. The reported error-bars are the standard errors over the set of trials.

Comparison With Least Squares Fitting Algorithms: Our aim in this experiment is to design a test that meaningfully demonstrates the overall effect of the structural component upon the recovery of affine projection parameters.



(a) Graph size 20 nodes.

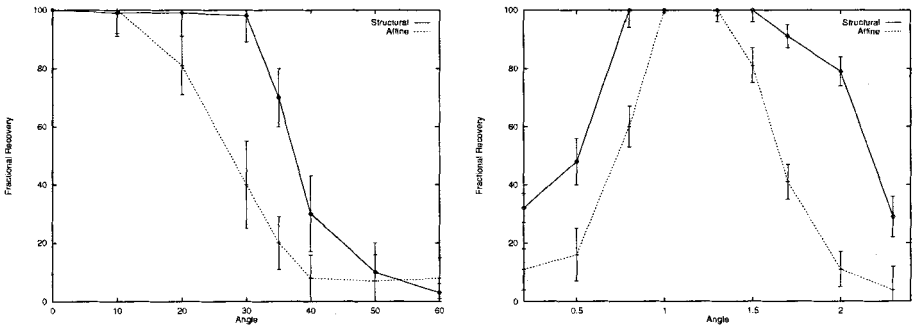
(b) Graph size 30 nodes.

Fig. 1. Comparing the holistic matching scheme with a least squares approach.

With this aim in mind, we have taken random point sets and used these as the model graphs in our experiments. We generate data-graphs by deleting a controlled number of points and re-inserting random new ones into the original model graphs. The fraction of modified points is taken as a measure of structural corruption. As a measure of success we have used average, unweighted, distance between point correspondences. This figure of merit includes only points that have a direct correspondence match. In other words, we exclude nodes deleted from the graphs in the re-triangulation step.

Figure 1a shows a comparison of the holistic matching scheme and the standard, unweighted, least squares method of affine parameter estimation. Here we show the average registered point-distance as a function of the fraction of correct correspondence matches. It is clear from this plot that the structural component plays a significant role in reducing the effect of outliers on the converged image registration. Even when only 10 of the 20 nodes are correctly matched, then the structural approach successfully recovers solutions that have an average point error of less than 0.01 with an insignificant standard-error. Figure 1b repeats this experiment but for point sets with 30 nodes. Once again, we note that even when the fraction of outliers is as high as 50%, then our holistic scheme manages to recover solutions with a much smaller average residual point-distance.

Comparison With Standard Structural Matching: In order to demonstrate the relative stability of the holistic matching scheme we will compare it with a structural graph matching scheme. The algorithm used in this comparison is essentially the discrete relaxation process of Wilson and Hancock [16]. This structural matching technique results solely from the iteration of the MAP update process defined in equation (13), leaving the parameter estimates static. The aim of our study is to demonstrate the sensitivity of our method to isotropic



(a) Affine rotation. (b) Affine scaling.

Fig. 2. Sensitivity study.

image scale and rotation about the axis normal to the image plane. We will again base our study on large samples of random dot patterns.

The results of these sensitivity comparisons for rotation and scale are shown in Figure 2a and 2b respectively. In these figures, the x-axis shows the initial rotation (Figure 2a) or scale (Figure 2b) difference between the images being matched. The y axis, on the other hand, represents the fractional number of nodes correctly matched upon convergence. The dotted lines show the sensitivity of the standard structural matching scheme, while the solid lines are for the holistic matching method. It is clear that our holistic expectation-maximisation approach performs better than the standard relational matching scheme. In particular, the range of both rotation and scale over which the EM scheme successfully recovers meaningful results is significantly greater than that for the purely structural scheme. For instance, our method copes well with angle differences of up to 35 degrees, whereas the structural method must be initialised to within 10 degrees. In the case of the scale difference, the holistic method copes with differences in the range 0.7 to 1.6, whereas the the MAP scheme only functions effectively over the range 0.9 to 1.1. However, it must be stressed that the structural method can be rendered considerably more robust if affine invariant measures are used to compute the initial *a posteriori* matching probabilities.

5.2 Real World Imagery

In order to demonstrate the effectiveness of the new matching process on real world imagery we will consider the following two data-sets:

- **Disk Set:** This data set consists of a set of digital photographs of 3.5 inch floppy disks. This data-set was chosen since it allows for controlled shifts in viewpoint to be made. Both small viewpoint shifts that are nearly affine, and very large shifts where the controlled introduction of strong perspective

foreshortening will be investigated. Experiments with this data are aimed at evaluating how our matching method degrades when the geometric transformation departs from the assumed affine model.

- **Road Network:** In this experiment we are concerned with the registration of aerial infra-red images against a digital map. The images were taken at night-time and the most prominent features are those that radiate absorbed heat. In the urban scenes under study these features are the tarmac roads. We therefore chose the road networks as the basis for our graph structures. The nodes in our graphs are junctions detected in the road network. It is important to note that these images are distorted due to the geometry of the line-scan process. The images are captured using horizontal line-scan as the aircraft moves in the vertical direction. The line-scan process is controlled by the rotation of a mirror. For this reason the images are subject to barrel distortion in the x-direction. In the y-direction there are also sampling irregularities due to the aircraft changing heading due to banking or turbulence.

We will first consider the task of recognising planer objects in different 3D poses, which is posed by the set of images of floppy disks. The object used in this study is placed on a desktop. The different object viewing angles are contrived so as to introduce increasing degrees of perspective foreshortening. The feature points used to triangulate the object are corners which are extracted by hand. Figure 3 shows a sequence of object-views with the triangulations of the hand segmented feature-points superimposed. The first oblique view in the sequence is taken as the object-model; the remaining object-poses are used to test the matching process.

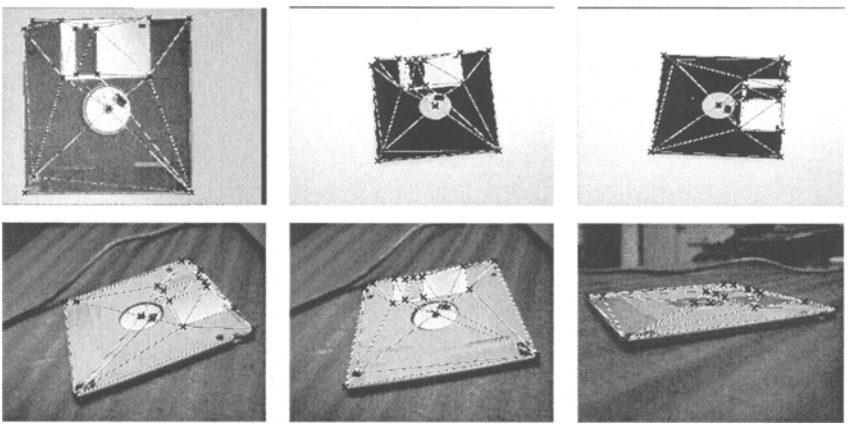


Fig. 3. The six views of used in the matching experiments.

Figure 4 shows the initial and final poses for the registration of the first and second images in the dataset. The fraction of correct initial correspondences

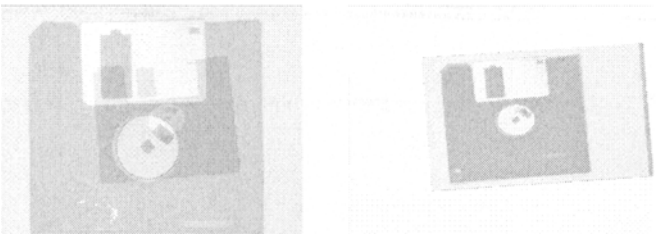


Fig. 4. a) An initial guess b) The final registration.

was found to be 50 percent. From the superimposed images it is clear that the recovered pose is accurate. Moreover, the ratio of the residual point registration error to the linear image dimension is 0.029.

Next, we provide an illustration of the iterative properies of our matching algorithm when one of the images under match exhibits a degree of perspective foreshortening. Clearly, in this case the affine transformation is no longer sufficient to represent the image deformation. Figure 5 shows the iterative registration for this experiment. The registration quickly converges upon a pose that is a good approximation to the full perspective transformation. In figure 6 we show the Delaunay triangulation iterating in synchronization with the image registration of figure 5. It is interesting to note the structure of the triangulation changing with iteration number. This clearly illustrates the effectivenness of the graph edit process in controlling the topology of the graph in the registration sequence.

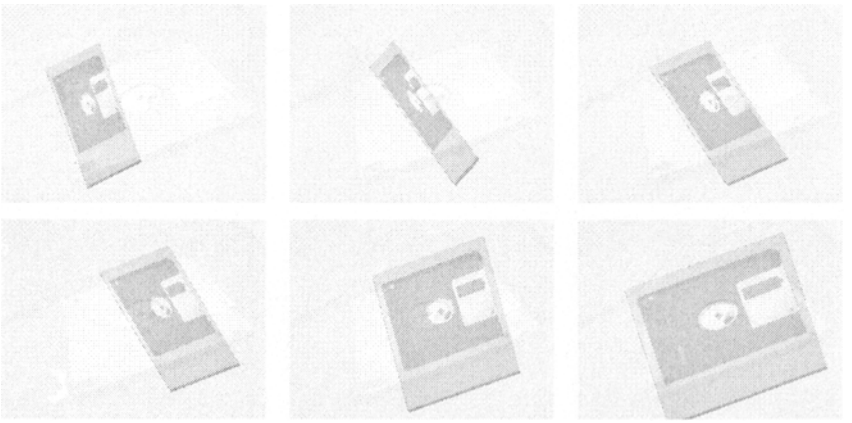


Fig. 5. Iterative convergence using an affine transformation.

The final piece of experimentation involves the registration of a digital map against a set of aerial infra-red images. Figure 7 shows the map data together

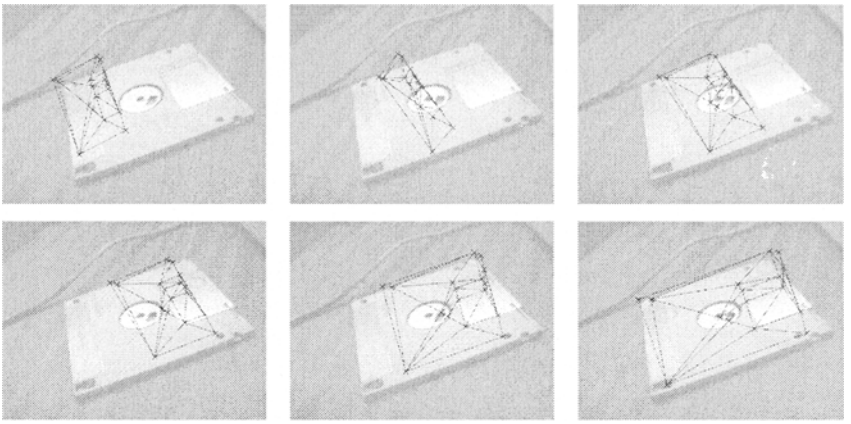


Fig. 6. The graphs iterating in synchronisation with the registration process.

Fig. 7. Aerial image registration: a) the digitial map; b) the registration with the high altitude image; c) the registration with the low altitude image.

with the raw images used in this example. The salient structure in this imagery is a road network. The feature points used in our matching experiments are junctions in the road network. These points are used to seed the Delaunay triangulation. There are three factors which complicate the matching process. Firstly, there are cartographic errors. As a result, there are features for which no correspondence exists even when the map is brought into exact registration with the images. Secondly, there is a significant amount of barrel distortion in these images. This process is not faithfully captured by our affine transformation model. Finally, the extracted Delaunay triangulations exhibit a significant degree of structural corruption. Figure 7 shows the final affine transformations of the map superimposed on the different aerial images. The matching process commences from a random initial estimate of the affine transformation matrix. It is clear that the recovered transformations are reasonably accurate given the poor geometric model.

6 Conclusions

Our main contribution in this paper has been to develop a new holistic matching algorithm. This two-step iterative process involves coupled operations to locate point-correspondences and estimate geometric transformation parameters. Point correspondences are located by maximum *a posteriori* graph-matching. Maximum likelihood parameters are recovered using the expectation-maximisation algorithm. These coupled iterative processes communicate by exchanging separate pieces of matching information. The point-correspondences passed by the matching process improve the robustness of maximum likelihood parameter estimation. In their turn, the maximum likelihood parameters are used to estimate *a posteriori* measurement probabilities which improve the accuracy of the point-correspondences.

We illustrate the effectiveness of the resulting matching process under affine geometry. This is a task of generic importance in computer vision with application in image mosaicking, pose recovery and camera calibration. Here the coupled matching process is shown to outperform structural matching. Moreover, the use of point-correspondences is shown to offer significant advantages in the control of added image noise.

In other words, we have presented a flexible matching method which unifies relational graph matching and pose-recovery. The framework is Bayesian and relies on some fairly non-restrictive assumptions concerning the Gaussian origin of measurement errors and observational independence. Our future plans revolve around the use of improved optimisation methods and more ambitious point-deformation models.

References

1. Y. Amit and A. Kong, "Graphical Templates for Model Registration", *IEEE PAMI*, **18,**, pp. 225–236, 1996.
2. A.D.J. Cross and E.R. Hancock, "Perspective Pose Recovery with a Dual Step EM Algorithm", *Advances in Neural Information Processing Systems*, **10**, Edited by M. Jordan, M. Kearns and S. Solla, MIT Press, 1998.
3. A.P. Dempster, 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. O.D. Faugeras, E. Le Bras-Mehlman and J-D. Boissonnat, "Representing Stereo Data with the Delaunay Triangulation", *Artificial Intelligence*, **44**, pp. 41–87, 1990.
5. Gold S., Rangarajan A. and Mjolsness E., "Learning with pre-knowledge: Clustering with point and graph-matching distance measures", *Neural Computation*, **8**, pp. 787–804, 1996.
6. R.I. Hartley, "Projective Reconstruction and Invariants from Multiple Images", *IEEE PAMI*, **16**, pp. 1036—1041, 1994.
7. M.I. Jordan and R.A. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation*, **6**, pp. 181-214, 1994.
8. M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Maalsburg, R.P. Wurtz and W.Konen, "Distortion-invariant object-recognition in a dynamic link architecture", *IEEE Transactions on Computers*, **42**, pp. 300–311, 1993

9. D.P. McReynolds and D.G. Lowe, "Rigidity Checking of 3D Point Correspondences under Perspective Projection", *IEEE PAMI*, **18** , pp. 1174–1185, 1996.
10. S. Moss and E.R. Hancock, "Registering Incomplete Radar Images with the EM Algorithm", *Image and Vision Computing*, **15**, 637–648, 1997.
11. D. Oberkampf, D.F. DeMenthon and L.S. Davis, "Iterative Pose Estimation using Coplanar Feature Points", *Computer Vision and Image Understanding*, **63**, pp. 495–511, 1996.
12. C. J. Poelman and T. Kanade, "A Para-perspective Factorization Method for Shape and Motion Recovery ", *IEEE PAMI*, **19**, 1997.
13. P. Torr, A. Zisserman and S.J. Maybank, "Robust Detection of Degenerate Configurations for the Fundamental Matrix", *Proceedings of the Fifth International Conference on Computer Vision*, pp. 1037–1042, 1995.
14. M. Tuceryan and T Chorzempa, "Relative Sensitivity of a Family of Closest Point Graphs in Computer Vision Applications", *Pattern Recognition*, **25**, pp. 361–373, 1991.
15. J. Utans, "Mixture Models and the EM Algorithms for Object Recognition within Compositional Hierarchies", *ICSI Berkeley Technical Report*, TR-93-004, 1993.
16. R.C. Wilson and E.R. Hancock, "Structural Matching by Discrete Relaxation", *IEEE PAMI*, **19**, pp. 634–648, 1997.