Short Papers

Correspondence Matching with Modal Clusters

Marco Carcassoni, *Member*, *IEEE*, and Edwin R. Hancock, *Member*, *IEEE*

Abstract—The modal correspondence method of Shapiro and Brady aims to match point-sets by comparing the eigenvectors of a pairwise point proximity matrix. Although elegant by means of its matrix representation, the method is notoriously susceptible to differences in the relational structure of the point-sets under consideration. In this paper, we demonstrate how the method can be rendered robust to structural differences by adopting a hierarchical approach. To do this, we place the modal matching problem in a probabilistic setting in which the correspondences between pairwise clusters can be used to constrain the individual point correspondences. We demonstrate the utility of the method on a number of synthetic and real-world point-pattern matching problems.

Index Terms—Point-pattern matching, spectral graph theory, robust statistics, hierarchy.

1 Introduction

THE graph spectral analysis of proximity data has proven to be an alluring yet elusive method for the tasks of correspondence matching and object recognition in computer vision. Stated simply, the aim is to find the pattern of correspondence matches between two sets of objects using the eigenvectors of an adjacency matrix or an attribute proximity matrix. The problem draws on spectral graph theory [3] and has been extensively studied for both the abstract problem of graph-matching [19], [16], and for point pattern matching [15], [14]. For instance, Umeyama [19] has developed an eigendecomposition method for exact graph-matching. In related work, Sossa and Horaud [17] have used spectral methods for the recognition of line-drawings using immanantal polynomials for the Laplacian matrix. There have also been a number of attempts to use spectral methods for point-set matching. Scott and Longuet-Higgins [14] align point-sets by performing singular value decomposition on a point association weight matrix. This method has recently been extended by Pilu [12] who includes neighborhood intensity correlation information into the association weight calculation. To overcome problems with the Scott and Longuet-Higgins method for large rotation angles, Shapiro and Brady [15] have reported a correspondence method which relies on measuring the similarity of the eigenvectors of a Gaussian point-proximity matrix. Provided that the point-sets are of the same size, then the correspondences delivered by the Shapiro and Brady method are relatively robust to random point jitter and to affine rotations and scaling. More recent work on the spectral analysis of point-sets includes that of Sclaroff and Pentland [13] and Cootes et al. [4], both of which aim to develop deformable models of shape.

One of the limitations with spectral methods is that they are particularly susceptible to the effect of size difference and structural error. In other words, spectral graph theory can furnish very efficient methods for characterizing exact relational structures, but soon breaks down when there are spurious nodes and edges in the graphs under study. In recent work, Luo and Hancock [9] have attempted to overcome these problems for the graph-matching problem. They have shown how the Umeyama algorithm can be rendered robust to size differences and structural differences in the edge-sets by using the statistical apparatus of the EM algorithm. It is also worth noting, that the EM algorithm has been used by a number of other authors for both rigid and nonrigid point-set matching. For instance, Cross and Hancock [5] show how relational constraints may be embedded in the algorithm and exploited to improve the accuracy of alignment and, hence, reduce correspondence errors. Chui and Rangarajan [2] have used deterministic annealing to control the certainty of the correspondence probabilities. However, neither of these two pieces of work uses graph-spectral information. Finally, in a recent paper [1], we have taken some steps toward improving the robustness of the Shapiro and Brady [15] method to point position movement, and have overcome the problems of structural error by resorting to an explicit alignment process based on the EM algorithm.

In this paper, we take this work one step further by focusing on how the correspondence process can be rendered robust to structural differences in the point-sets without the need for explicit alignment. We adopt a hierarchical approach to the correspondence problem. The method is based on the observation that the modes of the proximity matrix can be viewed as pairwise clusters. Rather than explicitly grouping the points prior to matching, here we aim to characterize the potential groupings in an implicit or probabilistic way and to exploit their arrangement to provide constraints on the pattern of correspondences. The hierarchy hence consists of two levels: At coarse detail, or high level, we perform modal analysis to find point cluster center correspondences. At fine detail, or low level, we use the cluster center correspondences to constrain the individual point correspondences. Hence, our method is only applicable to point sets which have a reasonably well-defined cluster structure. The probabilities used to model the correspondence process are chosen heuristically rather then being modeled from first principals using an error propagation analysis for the components of the eigenvectors. Although this represents a shortcoming, the empirical results obtained are encouraging.

2 Background

In this section, we review the Shapiro and Brady method for the modal matching of point-sets and detail the main conclusions of our own recent work aimed at improving the method through a better choice of proximity matrix weighting function and a more sophisticated means of comparing the modal coefficients [1].

We are interested in finding the correspondences between two point-sets, a *model* point-set $\mathbf{z}$, and a *data* point-set $\mathbf{w}$. Each point in the image *data* set is represented by a coordinate vector $\underline{\mathrm{w}}_i = (x_i, y_i)^T$, where i is the point index. In the interests of brevity, we will denote the entire set of image points by $\mathbf{w} = \{\underline{\mathrm{w}}_i, \forall i \in \mathcal{D}\}$, where $\mathcal{D}$ is the point index-set. The corresponding fiducial points constituting the *model* are similarly represented by $\mathbf{z} = \{\underline{\mathrm{z}}_j, \forall j \in \mathcal{M}\}$, where $\mathcal{M}$ denotes the index-set for the model feature-points $\underline{\mathrm{z}}_j$.

In Shapiro and Brady's original work, the weighting function was the Gaussian [15]. If i and i' are two data points, then the corresponding element of the proximity matrix is given by

$$H_D(i,i') = \exp\left[-\frac{1}{2s^2}||\vec{\boldsymbol{w}}_i^{(n)} - \vec{\boldsymbol{w}}_{i'}^{(n)}||^2\right], \quad (1)$$

where s is the width of kernel. However, in our recent paper, we have shown that alternatives to this weighting function suggested by the robust statistics literature offer better performance. In robust statistics, the weight (or influence function) $H(\eta)$ may be obtained from a potential function (or error kernel) $\rho(\eta)$ via the relationship $H(\eta) = \frac{1}{\eta}\frac{d\rho(\eta)}{d\eta}$, where η is the error residual. There are many concrete examples for the weight and potential functions, including Tukey's biweight, Huber's error kernel, and Li's adaptive

• *The authors are with the Department of Computer Science, University of York, York YO1 5DD, UK. E-mail: {marco, erh}@cs.york.ac.uk.*

Manuscript received 7 May 2002; revised 14 Jan. 2003; accepted 26 May 2003. Recommended for acceptance by Z. Zhang.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116488.

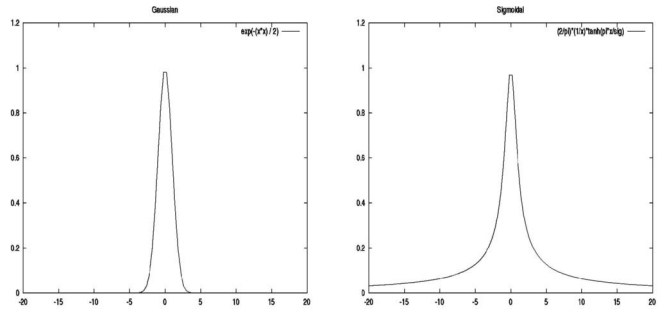

Fig. 1. Weighting functions: Gaussian and sigmoidal kernels.

penalty functions. However, one of the most effective and flexible alternatives is the weight function that results from Green's potential function $\rho(\eta) = \frac{s}{\eta}\log\cosh\frac{\pi\eta}{s}$ [7]. The function was used to generate a Markov prior for the purposes of Bayesian reconstruction of tomographic images. The $\log\cosh$ function was adopted by Green since it is flexible and can approximate both the Gaussian and the linear potential function for suitable choices of s; it may also be approximated by a piecewise polynomial and is hence related to the Huber kernel. Moreover, the potential function gives a convex energy. The weighting function corresponding to Green's potential is

$$H_D(i,i') = \frac{1}{||\vec{\boldsymbol{w}}_i^{(n)} - \vec{\boldsymbol{w}}_{i'}^{(n)}||}\tanh\left[\frac{\pi}{s}||\vec{\boldsymbol{w}}_i^{(n)} - \vec{\boldsymbol{w}}_{i'}^{(n)}||\right]. \tag{2}$$

The two weighting functions given in (1) and (2) are compared in Fig. 1. One of the conclusions of our recent study was that the tanh function gives a significant improvement in the fraction of correct correspondences when the point-set is subjected to Gaussian position error. However, both give poorer performance under point position jitter than a weighting function that is either piecewise Euclidean or a reciprocal (i.e., one for which the derivative of the potential function is increasing). For structural error, the tanh is better than both of these alternatives and is also easier to control.

The second contribution in our recent study [1] was to suggest an improved means of using the eigenvectors of the proximity matrix to compute correspondences for point-sets that are subject to jitter but are free from structural error. Consider the proximity matrix H_D for the set of data-points. Suppose that λ_l^D is the lth eigenvalue of the matrix H_D and that ϕ_l^D is the corresponding eigenvector. The modal structure of the point-sets is found by solving the eigenvalue equation $|H_D - \lambda I| = 0$ and the associated eigenvector equation $H_D\phi_l^D = \lambda_l^D\phi_l^D$.

Suppose that the suffix of the eigenvectors refers to the magnitude order of the eigenvalues, i.e., $\lambda_1^D \geq \lambda_2^D \geq \lambda_3^D \geq \ldots$, etc. We concatenate the eigenvectors in this order to construct a modal matrix

$$\Phi_D = \left(\phi_1^D|\phi_2^D|\phi_3^D|\ldots\right).$$

The column index of this matrix refers to the order of the eigenvalues while the row-index is the index of the original point-set. This modal decomposition is repeated for the model point-set to give a model-point modal matrix

$$\Phi_M = \left(\phi_1^M|\phi_2^M|\ldots|\phi_{|\mathcal{M}|}^M\right).$$

Since the two point-sets are potentially of different size, we truncate the modes of the larger point-set. This corresponds to removing the last $||\mathcal{D}| - |\mathcal{M}||$ rows and columns of the larger matrix. The resulting matrix has $o = \min\{|\mathcal{D}|, |\mathcal{M}|\}$ rows and columns.

Based on the coefficients of the modal matrices Φ_M and Φ_D, we aim to find correspondences between points. To this, we compute the probability $\zeta_{i,j}$ that the node $i \in \mathcal{D}$ matches to the node $j \in \mathcal{M}$. The node i is placed in correspondence with node j if $j = \arg\max_{j^*}\{\zeta_{i,j^*}\}$. In the remainder of this section, we discuss how the correspondence probabilities $\zeta_{i,j}$ may be computed from the coefficients of the modal matrices Φ_D and Φ_M.

Shapiro and Brady [15] find correspondences that minimize the Euclidean distance between the rows of the model matrices Φ_M and Φ_D. The correspondence probabilities are assigned according to the binary rule

$$\zeta_{i,j} = \begin{cases} 1 & \text{if } j = \arg\min_{j'}\sum_{l=1}^{o}||\Phi_D(i,l) - \Phi_M(j',l)||^2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

To render the computation of correspondences robust to outlier measurement error, we adopted an approach in which the probabilities are computed on a component by component basis over the eigenvectors so that large component differences contribute insignificantly. The probability that node i is in correspondence with node j is

$$\zeta_{i,j} = \frac{\sum_{l=1}^{o}\exp\left[-\mu||\Phi_D(i,l) - \Phi_M(j,l)||^2\right]}{\sum_{j'\in\mathcal{M}}\sum_{l=1}^{o}\exp\left[-\mu||\Phi_D(i,l) - \Phi_M(j',l)||^2\right]}. \tag{4}$$

3 Modal Clusters

These two refinements of the Shapiro and Brady method can render the correspondence process robust to significant movements in the point positions, but do not overcome problems associated with the different size of the point-sets. However, the correspondence probabilities proved useful for performing point-set alignment using the EM algorithm [1]. The aim in this paper, on the other hand, is to render the modal correspondence process robust to differences in point-set size due to drop-out, occlusion, or contamination. To achieve this goal, we adopt a hierarchical approach. The idea is to use the eigenmodes of the proximity matrix to identify pairwise clusters. Correspondence probabilities computed from the cluster center proximity matrices are used to constrain the individual point correspondences. Our clustering process is based on an analysis of the eigenmodes of the point proximity matrix. We use the coefficients of the first S columns of the modal matrix Φ_D to define the clusters. The set of points belonging to the cluster indexed ω_d is

$$\mathcal{C}_{\omega_d}^D = \left\{i|\frac{|\Phi_D(i,\omega_d)|}{\sum_{l=1}^{S}|\Phi_D(i,l)|} > T\right\}, \tag{5}$$

where $\omega_d = 1,\ldots S$ and T is a cluster membership threshold which, in practice, is set to be 0.95. Although we do not have room to discus the sensitivity of the method to the choice of this parameter, provided that it is set close to unity, the choice does not unduly affect the performance of the correspondence method. The number of clusters S may be chosen using the cumulative eigenvalue ratio

$$R_S = \frac{\sum_{i=1}^{S}\lambda_i^2}{\sum_{i=1}^{o}\lambda_i^2}.$$

We choose S so that $R_S > 0.95$, i.e., the clusters capture 95 percent of the point-set position variance.

For the cluster indexed ω_d, the position-vector for the cluster center is

$$\underline{\mathbf{c}}_{\omega_d}^{D(n)} = \frac{\sum_{i=1}^{|\mathcal{D}|}|\Phi_D^{(n)}(i,\omega_d)|\underline{\mathbf{w}}_i}{\sum_{i=1}^{|\mathcal{D}|}|\Phi_D^{(n)}(i,\omega_d)|}. \tag{6}$$

From the cluster center positions associated with the S largest eigenvalues, i.e., the first S columns of Φ_D, we use the robust weighting kernel to compute an $S \times S$ cluster center proximity matrix

$$G_D^{(n)}(l,l') = \frac{1}{||\underline{c}_l^{D(n)} - \underline{c}_{l'}^{D(n)}||}\tanh\left[\frac{\pi}{s}||\underline{c}_l^{D(n)} - \underline{c}_{l'}^{D(n)}||\right]. \quad (7)$$

Our idea is to use the modes of the $S \times S$ cluster-center proximity matrix G_D for the purposes of matching. Accordingly, we solve the equation $det(G_D - \Lambda^D I) = 0$ to locate the eigenvalues of the modal or cluster-center proximity matrix. The eigenvectors ψ_l, $l = 1, \ldots, S$ of the cluster-center proximity matrix are found by solving the equation $G_D\psi_l^D = \Lambda_l^D\psi_l^D$. As before, these eigenvectors can be used to construct a modal-matrix for the cluster center positions. The matrix has the eigenvectors of G_D as columns, i.e.,

$$\Psi_D = \left(\psi_1^D|\psi_2^D|\ldots|\psi_S^D\right).$$

This procedure is repeated to construct a second $S \times S$ cluster-center modal matrix Ψ_M for the set of model points z. Since the principal modal-clusters are selected on the magnitude-order of the associated eigenvalues, there is no need to reorder them.

For the points belonging to each cluster, we also construct a within-cluster proximity matrix. To construct this matrix, we will need to relabel the points using a cluster point index which runs from 1 to $|\mathcal{C}_{\omega_d}|$. Accordingly, we let δ_{i,ω_d}^D denote the point-index assigned to the node i in the cluster ω_d. The proximity matrix for the points belonging to this cluster is denoted by F_{ω_D} and the corresponding modal matrix is $\Theta_{\omega_d}^D$. The modal matrix for the cluster indexed ω_m in the model point-set is denoted by $\Theta_{\omega_m}^M$.

4 Matching

The aim in this paper is to explore whether the additional information provided by the modal clusters can be used to improve the robustness of the matching process to point addition and dropout. We would like to compute the probability $P(i \leftrightarrow j)$, that the data-point $i \in \mathcal{D}$ is in correspondence with the model data-point $j \in \mathcal{M}$. To do this, we construct a *mixture model* over the set of possible correspondences between the set of S modal clusters extracted from the data point positions and the model point positions. Suppose that ω_d and ω_m, respectively, represent labels assigned to the modal clusters of the data and model point-sets. Applying the Bayes formula, we can write

$$P(i \leftrightarrow j) = \sum_{\omega_d=1}^{S}\sum_{\omega_m=1}^{S} P(i \leftrightarrow j|\omega_d \leftrightarrow \omega_m)P(\omega_d \leftrightarrow \omega_m), \quad (8)$$

where $P(i \leftrightarrow j|\omega_d \leftrightarrow \omega_m)$ represents the cluster-conditional probability that the node i belonging to the data-graph cluster ω_d is in correspondence with the node j that belongs to the model-graph cluster ω_m. The quantity $P(\omega_d \leftrightarrow \omega_m)$ denotes the probability that the data point-set cluster indexed ω_d is in correspondence with the model point-set cluster indexed ω_m.

4.1 Cluster Conditional Correspondence Probabilities

To compute the cluster-conditional point correspondence probabilities, we use the modal structure of the within-cluster proximity matrices. These correspondence probabilities are computed using the method outlined in (4) since, as we will see later, this proves to be the most effective of the alternatives. As a result, we write

$$P(i \leftrightarrow j|\omega_d \leftrightarrow \omega_m) = \frac{\sum_{l=1}^{O_{\omega_d,\omega_m}} \exp\left[-k_w||\Theta_{\omega_d}^D(\delta_{i,\omega_d}^D, l) - \Theta_{\omega_m}^M(\delta_{j\omega_m}^D, l)||^2\right]}{\sum_{j' \in \mathcal{M}}\sum_{l=1}^{O_{\omega_d,\omega_m}} \exp\left[-k_w||\Theta_{\omega_d}^D(\delta_{i,\omega_d}^D, l) - \Theta_{\omega_m}^M(\delta_{j',\omega_m}^M, l)||^2\right]}, \quad (9)$$

where $O_{\omega_d,\omega_m} = min[|\mathcal{C}_{\omega_m}|, |\mathcal{C}_{\omega_d}|]$ is the size of the smaller cluster.

4.2 Cluster Correspondence Probabilities

We have investigated two methods for computing the cluster correspondence probabilities $P(\omega_d \leftrightarrow \omega_m)$:

- **Modal eigenvalues:** The first method used to compute the cluster-center correspondence probabilities relies on the similarity of the normalized eigenvalues of the cluster-center modal matrix. The probabilities are computed in the following manner

$$P(\omega_d \leftrightarrow \omega_m) = \frac{\exp\left[-k_e\left\{\frac{|\Lambda_{\omega_d}^D|}{\sum_{\omega_d=1}^{S}|\Lambda_{\omega_d}^D|} - \frac{|\Lambda_{\omega_m}^M|}{\sum_{\omega_m=1}^{S}|\Lambda_{\omega_m}^M|}\right\}^2\right]}{\sum_{\omega_m=1}^{S}\exp\left[-k_e\left\{\frac{|\Lambda_{\omega_d}^D|}{\sum_{\omega_d=1}^{S}|\Lambda_{\omega_d}^D|} - \frac{|\Lambda_{\omega_m}^M|}{\sum_{\omega_m=1}^{S}|\Lambda_{\omega_m}^M|}\right\}^2\right]}. \quad (10)$$

- **Modal coefficients:** The cluster center correspondence probabilities have also been computed by performing a robust comparison of the coefficients of the modal matrices of the cluster-center proximity matrix. This is simply an application of the method outlined in (4) to the modal coefficients of the between-cluster proximity matrix. We therefore set

$$P(\omega_d \leftrightarrow \omega_m) = \frac{\sum_{L=1}^{S}\exp\left[-k_b||\Psi_D(\omega_d, L)| - |\Psi_M(\omega_m, L)||^2\right]}{\sum_{\omega_m=1}^{S}\sum_{L=1}^{S}\exp\left[-k_b||\Psi_D(\omega_d, L)| - |\Psi_M(\omega_m, L)||^2\right]}. \quad (11)$$

In the above equations, k_m, k_b, and k_e are exponential constants. Again, the choice of the value is not critical to the performance of the method, and setting all three constants to 0.1 gave good results in our experiments. There is clearly scope for developing a means of estimating these three parameters from the statistics of the interpoint distances and the intercluster distances.

5 Experiments

In this section, we describe our experimental evaluation of the new modal correspondence method. We commence with a sensitivity study in which we compare the new correspondence method with that of Shapiro and Brady [15] and the EM alignment method outlined in our recent paper [1]. The Shapiro and Brady method is based purely on modal correspondence analysis, while the alignment method uses modal correspondence probabilities to weight the estimation of affine alignment parameters in a dual-step EM algorithm.

Our sensitivity study uses randomly generated point-sets. We ensure that the point-sets have a clump structure by sampling the point positions from six partially overlapping Gaussian distributions with controlled variance. We have then added both new points at random positions and random point-jitter to the synthetic data. The randomly inserted points have been sampled from a uniform distribution. The positional jitter has been generated by displacing the points from their original positions by Gaussian measurement errors. The displacements have been randomly sampled from a circularly symmetric Gaussian distribution of zero mean and controlled standard deviation.

We commence by investigating the way in which the elements of the point proximity matrix are calculated. The aim is to determine which of the weighting functions returns correspondences which are the most robust to point-position jitter alone. Fig. 2 shows the fraction of correct correspondences as a function of the standard deviation of the added Gaussian position errors. The standard deviation is recorded as a fraction of the average closest point distance. We compare the results obtained using the

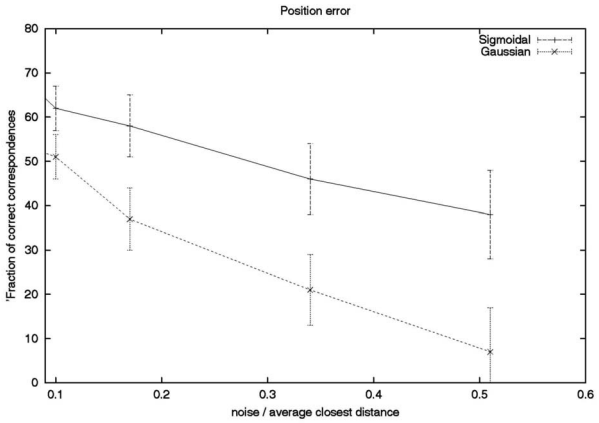

Fig. 2. Effect of weighting function on spectral correspondence in presence of position error.

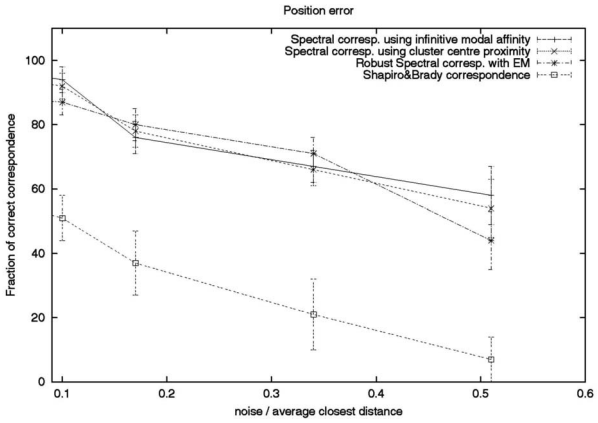

Fig. 4. Effect of spectral correspondence methods in presence of position error.

Gaussian weighting function (1) used by Shapiro and Brady and the use of the tanh function given (2). The tanh weighting function significantly outperforms the Gaussian weighting function. The margin of improvement is about 20 percent.

In Fig. 3, we show the effect of increasing the number of randomly added points. In this experiment, we commence with a point-set of size 100. The plot shows the fraction of points correctly matched as a function of the number of randomly added points. The long-dashed curve, i.e., the one which gives the consistently poorest performance, is the result of applying the Shapiro and Brady algorithm. Here, the fraction of correct correspondences falls below 25 percent once the fraction of added clutter exceeds 2 percent. The results obtained with the EM alignment method described in [1] are shown as a line-dot curve. This method performs best of all when the level of clutter is less than 20 percent. The remaining two curves show the results obtained with the two variants of our hierarchical correspondence algorithm detailed in Section 4. In the case of the dotted curve, the cluster correspondences are computed using only the modal coefficients of the between-cluster proximity matrix as described in (11). The solid curve shows the results obtained if the eigenvalues are also used as described in (10). There is little to distinguish the two methods. Both perform rather more poorly than the dual-step EM algorithm when the level of clutter is less than 20 percent. However, for larger clutter levels, they provide significantly better performance. The additional use of the eigenvalues results in a slight improvement in performance. Hence, the method reported in this paper outperforms the EM algorithm reported in our previous work at high levels of structural error.

Fig. 4 investigates the effect of positional jitter. Here, we plot the fraction of correct correspondence matches as a function of the standard deviation of the Gaussian position error added to the point-positions. We report the level of jitter using the ratio of the standard deviation of the Gaussian error distribution to the average closest interpoint distance. Here, there is nothing to distinguish the behavior of our hierarchical correspondence method from the dual-step alignment method. In each case, the fraction of correct correspondences degrades slowly with increasing point-position jitter. However, even when the standard deviation of the position errors is 50 percent of the average minimum interpoint-distance, then the fraction of correct correspondences is still greater than 50 percent. By contrast, the accuracy of the Shapiro and Brady method falls below 50 percent once the standard deviation of the positional error exceeds 10 percent of the minimum interpoint distance.

Our final set of experiments on synthetic data investigate the effect of diluting the cluster-structure of the point-sets. Here, we have gradually moved the cluster-centers closer together and have investigated the effect on the fraction of correct correspondences when there is structural error present. The results are shown in Fig. 5. Here, we show the fraction of correct correspondences as a function of the overlap between the clusters. We have also included tests to show the performance of the algorithm when 20 percent of clutter noise is added to the overlapping clusters. The results are shown as groups of bars. As we move from left to right across the plot, the degree of cluster overlap increases. In each group of bars, the leftmost (i.e., black) bar is the result obtained with the new method reported in this paper when the clusters are clutter-free, while the center (light gray) bar is the result obtained when 20 percent of clutter is added. The right-most bar (i.e., midgray) is the result obtained with the Shapiro and Brady method when the point-sets contain no clutter. It is important to stress that as the clusters are moved together, larger "superclusters" may develop due to merging. This effects the value of S used in our

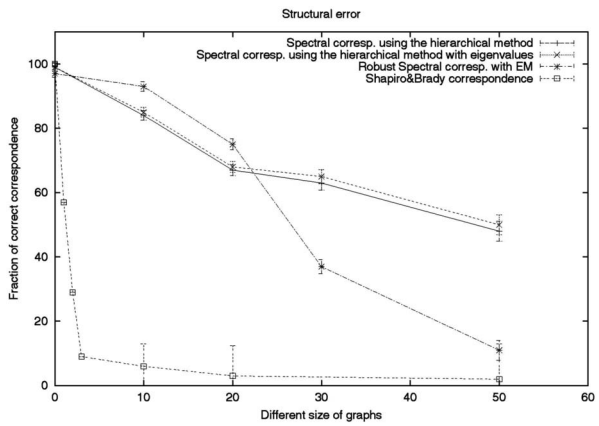

Fig. 3. Effect of spectral correspondence methods in presence of structural error.

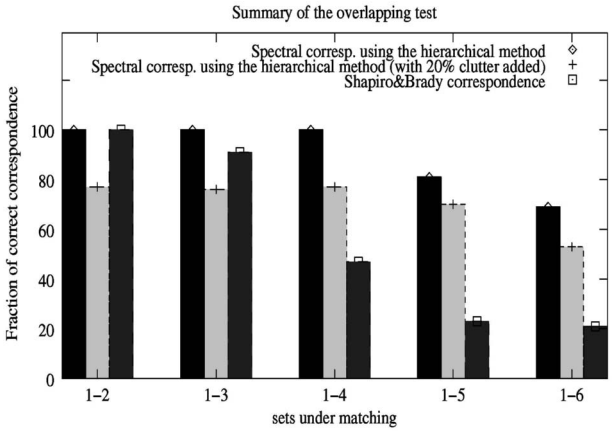

Fig. 5. Cluster stability.

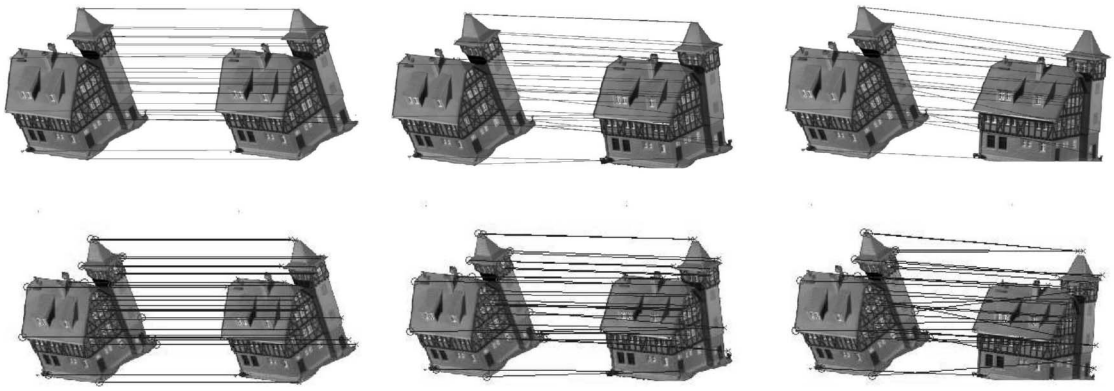

Fig. 6. Correspondences between the first image and subsequent images in the CMU house sequence using our new method (top row) and with the Scott and Longuett-Higgins matching algorithm (bottom row).

TABLE 1
Performance on the CMU/VASC House Sequence, where the House "0" is Tested against the Remaining Nine

# image	0	1	2	3	4	5	6	7	8	9
# of points	30	32	32	30	30	32	30	30	30	31
correctly matched	-	30	29	30	28	28	29	27	25	23
% matched	-	100%	97%	100%	94%	94%	97%	91%	83%	76%
correctly matched (SL-H)	-	30	29	29	29	28	25	21	17	15
% matched (SL-H)	-	100%	97%	97%	97%	94%	85%	73%	61%	55%

analysis and could, in principal, be detected using the quantity R_S. The performance of the Shapiro and Brady method is poorer than the new method. Its sudden drop-off in performance is attributable to the effect of increased point-density as the clusters are overlapped. Obviously, the performance of the new method degrades with the addition of clutter. However, the increased proximity of the clusters does not appear to significantly degrade performance.

We now turn our attention to real world data. Based on the sensitivity study, we confine our attention to the case where 1) the proximity matrix is as defined in (2), 2) the cluster center calculation is as outlined in Section 3, and 3) the modal coefficient probabilities are as described in (10). In the majority of our experiments, we are concerned with matching corner-features. We use the corner detector recently reported by Luo et al. [8] to extract point features. Our first experiments are performed with the CMU/VASC model-house sequence. The data set consists of a series of views of a model house, collected as the viewing direction changes. We match the first image of the sequence to each of the subsequent nine frames. Some example results for the second, fifth, and eighth frames are shown in the top row of Fig. 6. Here, the lines between feature points represent correspondences. We have compared the number of correct correspondences with ground-truth obtained by hand labeling. Table 1 lists the fraction of correct correspondences. This varies between 100 percent for the closest pair of views to 76 percent for the most distant pair of views. We have compared the results obtained with our new method with those obtained using the Scott and Longuet-Higgins method. The correspondences are shown in the bottom row of Fig. 6 and the results are summarized in Table 1. Comparing the results of the two methods, it is clear that the Scott and Longuet-Higgins method gives poorer results when the difference in viewing angle is large.

We have repeated these experiments for a number of different images. The first of these involves images from a gesture sequence of a hand. The image used in this study is shown in Fig. 7. To extract feature-points from this data, we commence by running the Canny edge detector over the images to locate the boundary of the hand. From this edge data, point features have been detected using the corner detector of Mokhtarian and Suomela [10]. The raw points returned by this method are distributed relatively uniformly along the outer edge of the hand and are hence not suitable for cluster analysis. We have therefore pruned the feature points using a curvature criterion. We have removed all points for which the

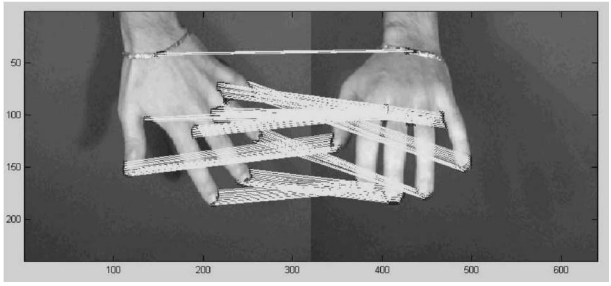

Fig. 7. Real data experimentation: Hierarchical correspondence method.

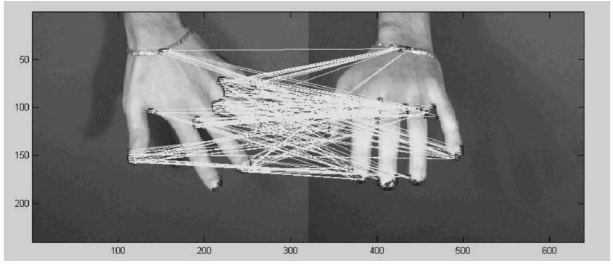

Fig. 8. Real data experimentation: Shapiro and Brady's correspondence method.

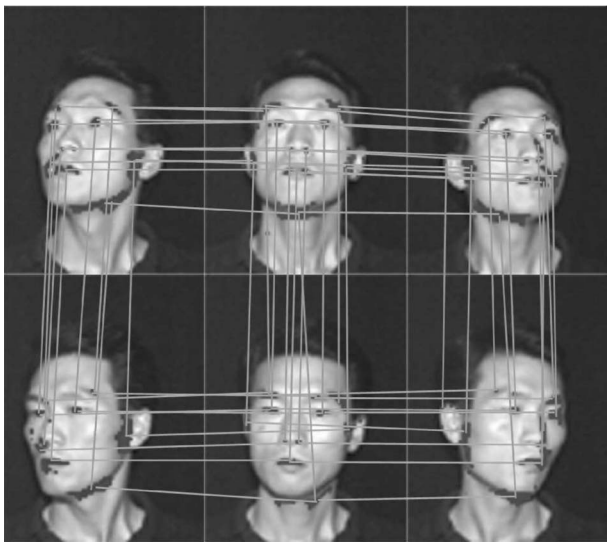

Fig. 9. Real data experimentation on tilting head: Hierarchical correspondence method.

curvature of the outline is smaller than a heuristically set threshold. Initially, there are some 800 feature points, but after pruning, this number is reduced to 271. The pruned feature-points are shown in blue in the figure. They are clustered around the fingertips and the points at which the fingers join the hand. After applying the clustering method, the set of centers shown in red is obtained. There are 10 clusters in both images. The yellow lines between the two images show the detected correspondences. The fraction of correct correspondences is 81.2 percent. Fig. 8 shows the correspondences obtained with the Scott and Longuet-Higgins method. The results are poorer, and the fraction of correct correspondences is 70 percent.

Our third real-world experiment involves a sequence of images obtained as a subject rotates and tilts his head. The feature points here are highly nonplanar. In Fig. 9, we show the correspondences obtained. These are again good, and there appear to be no systematic problems. Fig. 10 shows the results delivered by the Scott and Longuet-Higgins method. From the pattern of correspondence lines, it is clear that the method does not perform well. A fourth example is shown in Fig. 11 where we show the results obtained on an image pair from the roof-space of our lab. Here, the correspondences are good despite the fact that there is no clear

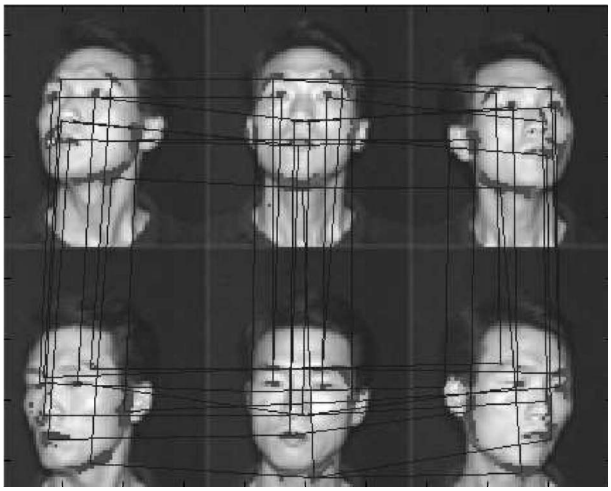

Fig. 10. Real data experimentation on tilting head: Shapiro and Brady's correspondence method.

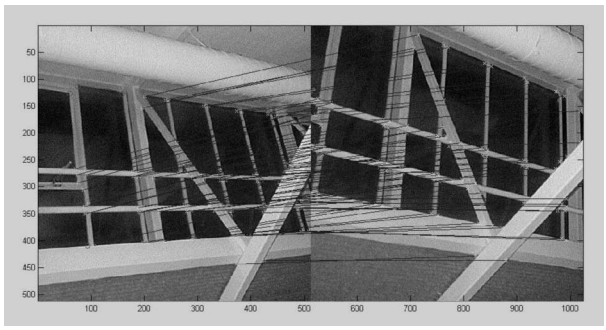

Fig. 11. Real data experimentation on roof-space images: Hierarchical correspondence method.

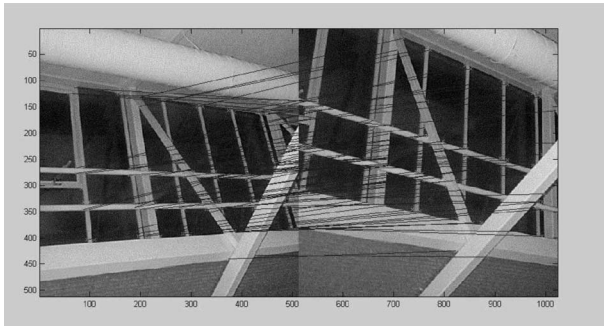

Fig. 12. Real data experimentation on roof-space images: Shapiro and Brady's correspondence method.

cluster-structure. The feature sets contain a different number of points and some of them appear in one image and not in the other. The results represent an improvement over those obtained with the Scott and Longuet-Higgins method in Fig. 12.

6 Conclusions

In this paper, we have investigated how the correspondence method of Shapiro and Brady [15] may be improved by using the modal coefficients of the point-proximity matrix to establish the whereabouts of significant point groupings or clusters. We exploit these groupings to develop a hierarchical correspondence method. This is a two-step process: First, we use the spatial arrangements of the center-points of the most significant groups to compute a between-cluster proximity matrix. The modal coefficients of this between-cluster proximity matrix are used to establish correspondence probabilities between groups of points. Second, for each group of points, we compute a within-cluster proximity matrix. The modal coefficients of these within-cluster proximity matrices are used to establish heuristic cluster-conditional point correspondence probabilities. Using the Bayes rule, we combine these two sets of probabilities to compute individual point correspondence probabilities. We have shown that, while the Shapiro and Brady method fails once more than a few percent of clutter is added, the new method degrades more gracefully.

Although we have concentrated on point pattern matching in this paper, some of the ideas presented are of more generic usefulness. In particular, the idea of using modal or spectral clusters to overcome problems of size difference may also prove useful for the problem of inexact graph-matching [6], [20], [21], [11]. Hence, we may have a route to improving the robustness of the Umeyama [19] algorithm. Moreover, the methodology reported in this paper may prove useful in improving the robustness of SVD-based methods for motion analysis [18] and stereopsis [12] to structural difference.

REFERENCES

[1] M. Carcassoni and E.R. Hancock, "Spectral Correspondence for Point Pattern Matching," *Pattern Recognition,* vol. 36, no. 1, pp. 193-204, 2003.
[2] H. Chui and A. Rangarajan, "Non-Rigid Point Matching Using Mixture Models," *Proc. Workshop Math. Methods in Biomedical Image Analysis,* pp. 190-197, 2000.
[3] F.R.K. Chung, *Spectral Graph Theory,* CBMS series, AMS ed. vol. 92, 1997.
[4] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models: Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 38-59, 1995.
[5] A.D.J. Cross and E.R. Hancock, "Graph Matching with Dual Step Em Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, pp. 1236-1253, 1998.
[6] S. Gold and A. Rangarajan, "A Graduated Assignment Algorithm for Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 4, pp. 377-388, Apr. 1996.
[7] P.J. Green, "Bayesian Reconstruction from Emission Tomography Data Using a Modified Em Algorithm," *IEEE Trans. Medical Imaging,* vol. 9, pp. 84-93, 1990.
[8] B. Luo, A.D.J. Cross, and E.R. Hancock, "Corner Detection via Topographic Analysis of Vector-Potential," *Pattern Recognition Letters,* pp. 635-650, 1999.
[9] B. Luo and E.R. Hancock, "Structural Matching Using the Em Algorithm and Singular Value Decomposition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, pp. 1120-1136, 2001.
[10] F. Mokhtarian and R. Suomela, "Robust Image Corner Detection through Curvature Scale Space," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, pp. 1376-1381, Dec. 1998.
[11] M. Pelillo, K. Siddiqi, and S.W. Zucker, "Matching Hierarchical Structures Using Association Graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 11, pp. 1105-1120, Nov. 1999.
[12] M. Pilu, "A Direct Method for Stereo Correspondence Based on Singular Value Decomposition," *Proc. IEEE Computer Vision and Pattern Recognition Conf.,* pp. 261-266, 1997.
[13] S. Sclaroff and A.P. Pentland, "Modal Matching for Correspondence and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 6, pp. 545-561, 1995.
[14] G.L. Scott and H.C. Longuett-Higgins, "An Algorithm for Associating the Features of Two Images," *Proc. Royal Soc. London B,* vol. 244, pp. 21-26, 1991.
[15] L.S. Shapiro and J.M. Brady, "Feature-Based Correspondence: An Eigenvector Approach," *Image and Vision Computing,* vol. 10, pp. 283-288, 1992.
[16] A. Shokoufandeh, S.J. Dickinson, K. Siddiqi, and S.W. Zucker, "Indexing Using a Spectral Encoding of Topological Structure," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 491-497, 1999.
[17] H. Sossa and R. Horaud, "Model Indexing: The Graph-Hashing Approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 811-815, 1992.
[18] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorisation Method," *Int'l J. Computer Vision,* vol. 9, no. 2, pp. 137-154, 1992.
[19] S. Umeyama, "An Eigen Decomposition Approach to Weighted Graph Matching Problems," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, pp. 695-703, 1988.
[20] R.C. Wilson, A.D.J. Cross, and E.R. Hancock, "Structural Matching with Active Triangulation," *Proc. Computer Vision and Image Understanding,* vol. 72, no. 1, pp. 21-38, Oct. 1998.
[21] R.C. Wilson and E.R. Hancock, "Structural Matching by Discrete Relaxation," *IEEE Trans. Pattern Recognition and Machine Analysis,* vol. 19, pp. 634-648, 1997.

The CMU Pose, Illumination, and Expression Database

Terence Sim, *Member*, *IEEE*,
Simon Baker, and Maan Bsat

Abstract—In the Fall of 2000, we collected a database of more than 40,000 facial images of 68 people. Using the Carnegie Mellon University 3D Room, we imaged each person across 13 different poses, under 43 different illumination conditions, and with four different expressions. We call this the CMU Pose, Illumination, and Expression (PIE) database. We describe the imaging hardware, the collection procedure, the organization of the images, several possible uses, and how to obtain the database.

Index Terms—Face databases, pose, illumination, expression.

1 INTRODUCTION

PEOPLE look very different depending on a number of factors. Perhaps the three most significant factors are: 1) the pose, i.e., the angle at which you look at them, 2) the illumination conditions at the time, and 3) their facial expression, e.g., smiling, frowning, etc. Although several other face databases exist with a large number of subjects [8], and with significant pose and illumination variation [2], we felt that there was still a need for a database consisting of a fairly large number of subjects, each imaged a large number of times, from several different poses, under significant illumination variation, and with a variety of expressions.

Between October 2000 and December 2000, we collected such a database consisting of more than 40,000 images of 68 subjects. (The total size of the database is about 40 GB.) We call this the CMU Pose, Illumination, and Expression (PIE) database. To obtain a wide variation across pose, we used 13 cameras in the CMU 3D Room [7]. To obtain significant illumination variation, we augmented the 3D Room with a "flash system" similar to the one constructed by Georghiades et al. at Yale University [2]. We built a similar system with 21 flashes. Since we captured images with and without background lighting, we obtained $21 \times 2 + 1 = 43$ different illumination conditions. Furthermore, we asked the subjects to pose with several different expressions.

In the remainder of this paper, we describe the capture hardware, the organization of the images, a large number of possible uses of the database, and how to obtain a copy of it.

2 CAPTURE APPARATUS

2.1 Setup of the Cameras: Pose

Obtaining images of a person from multiple poses requires either multiple cameras capturing images simultaneously, or multiple "shots" taken consecutively (or a combination of the two.) There are a number of advantages of using multiple cameras: 1) the

- *T. Sim is with the Department of Computer Science, National University of Singapore, 3 Science Drive 2, Singapore, 117543, Republic of Singapore. E-mail: tsim@comp.nus.edu.sg.*
- *S. Baker is with the Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: simonb@cs.cmu.edu.*
- *M. Bsat is with the Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: mbsat@andrew.cmu.edu.*

Manuscript received 20 Feb. 2002; revised 26 Sept. 2002; accepted 9 Apr. 2003.
Recommended for acceptance by P. Belhumeur.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 115919.