

# Kernel Spectral Correspondence Matching Using Label Consistency Constraints

Hongfang Wang and Edwin R. Hancock

Dept. of Computer Science, University of York,  
Heslington, York, YO10 5DD, UK  
{hongfang,erh}@cs.york.ac.uk

**Abstract.** This paper investigates a kernel spectral approach to the problem of point pattern matching. Our first contribution is to show how kernel principal components analysis can be effectively used for solving the point correspondence matching problem when the point-sets are subject to structural errors, i.e. they are of different size. Our second contribution is to show how label consistency constraints can be incorporated into the construction of the Gram matrices for solving the articulated point pattern matching problem. We compare our algorithm with earlier point matching approaches and provide experiments on both synthetic data and real world data.

## 1 Introduction

The problem of point pattern matching is to find one-to-one correspondences between two given point-sets and serves as an important component of many computer vision tasks. Graph spectral methods [4] have been used extensively for locating correspondences between feature point-sets, e.g. [13,14,10]. Scott and Longuet-Higgins [13] used a Gaussian weighting function to build an inter-image proximity matrix between feature points and used singular value decomposition (SVD) to locate correspondences. This method fails when the rotation or scaling between the two images being matched is too large. To overcome this problem, Pilu [10] introduces a *feature similarity measure* into the algorithm by incorporating a neighbourhood correlation measure into the proximity matrix. Shapiro and Brady [14] extend the Scott and Longuet-Higgins method and show how correspondences can be located using the eigenvectors of intra-image proximity matrices. Carcassoni and Hancock improve the robustness of the Shapiro and Brady method by using robust error kernels instead of the Gaussian weighting function [1], and exploit image structure using spectral clusters [2].

The location of correspondences between feature points belonging to non-rigid objects is a not only more challenging, but also a potentially more important task. Many existing approaches rely on such information. Examples include the point distribution model (PDM) of Cootes and Taylor [5], and the factorisation method of Tomasi and Kanade [15]. In the literature, many attempts have been described to recover accurate correspondences for non-rigid motion. For example, in [8] the softassign method is used to compute correspondences in a manner that

is robust to outliers. In [3], a thin-plate spline is used to model the non-rigid motion of curves and proves successful for point pattern matching.

An interesting source of information that can be used in non-rigid motion, but has received relatively little attention, is that provided by label consistency constraints. In many types of image, the points can be assigned semantic labels to distinguish their identity. Using this information the consistency of pairwise relations can be tested against a scene constraint model. Hence, correspondences which are inconsistent with the model can be rejected. In this paper, we aim to use label consistency information to construct a weighted kernel matrix, and to use this matrix to deliver more robust and computationally effective matching results. Our first contribution is to show how the point proximity matrix can be incorporated into the support function for relaxation labelling. In this way when the label probabilities are updated, then the strength of the proximity relations is brought to bear on the computation of label support. Our second contribution is to show how the label probabilities can be used to refine the point correspondence process using kernel PCA [12]. In our experiments we compare the performance of our algorithm with a number of previous approaches to point pattern matching. We demonstrate that with an appropriate choice of kernel function, the method delivers encouraging performance. In particular, the results are less sensitive to the problems that limit the performance of previous graph spectral methods.

## 2 Label Process

In the computer vision literature, one of the most extensively studied approaches to the consistent labeling problem involves the use of a discrete or continuous relaxation technique. In the continuous or probabilistic case, each node is assigned an initial weight or probability distribution. Iteratively, the label probabilities or weights are updated until a consistent distribution is reached. The performance of the labelling depends critically on the compatibility coefficients and the support function used to combine evidence in the iterative process. In [7], a dictionary is used, and in [9] the compatibility coefficients are represented as a vector which is learned offline. Here our compatibility model shares some properties in common with the compatibility vector in [9].

Consider the feature point-sets,  $\mathbf{y} = \{\mathbf{y}_j\}_{j=1}^n$ ,  $\mathbf{y}_j = (y_{j1}, y_{j2})$ , and  $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^m$ ,  $\mathbf{x}_i = (x_{i1}, x_{i2})$  that result from the motion of an articulated object. Here the former set is treated as the model point-set. We augment the feature vectors with a vector of label probabilities, which represent the likelihood of belonging to possible rigid components. Assume there are  $L$  labels (rigid components) in each feature point-set. Each image point  $\mathbf{x}_i$  can be assigned a label  $\theta_i \in \Omega$ , where  $\Omega = \{\omega_1, \dots, \omega_L\}$ . Denote by  $P(\theta_i = \lambda)$  the probability that node  $\mathbf{x}_i$  is labeled as  $\lambda$ ,  $\lambda \in \Omega$ . The vector  $\mathbf{p}_i = (P(\theta_i = \omega_1), \dots, P(\theta_i = \omega_L))^T$  represents the probability of assigning each of the possible labels to the point, with  $0 \leq P(\theta_i = \lambda) \leq 1$ , and  $\sum_{\lambda=1}^L P(\theta_i = \lambda) = 1$ . The matrix  $P$  with the label probability vectors as columns, i.e.,  $P = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)^T$ , represents the

label probability distribution over the entire point-set. Our ultimate aim is to locate correspondences between the two point-sets on the basis of the above-mentioned information by using spectral graph theory. Our label consistency model is derived from the model feature point-set, and the learned label compatibilities used to assign consistent point labels to the “data” point-set. First a label compatibility matrix  $R \in \mathbb{R}^{L \times L}$  is constructed so as to embody knowledge of the number of rigid components, i.e. labels, in each image, together with the semantic constraints that apply between each pair of object-labels. It has elements  $R_{ij} = 1$  if  $\mathbf{x}_i$  and  $\mathbf{x}_j$  come from the same rigid component, and is defined to be  $-1$  otherwise. This definition restricts the nodes to give total positive support to the nodes in the same group (i.e. rigid component) and to contribute a negative support to nodes outside the group. The proximity constraint is also acquired from the model image. Further, we assume that in any two consecutive image frames, the relative position of the rigid components of the object under study will not change dramatically. The label probabilities for the data point-set are updated iteratively commencing from a set of initial values. Updating is effected using neighbourhood support. Let us denote the neighbourhood for the point  $\mathbf{x}_i$  and its  $k$  closest points by  $N_i = \{\mathbf{x}_{i1}, \dots, \mathbf{x}_{ik}\}$ . The support from the neighbourhood for the label assignment  $\lambda_i$  to point  $\mathbf{x}_i$  is:

$$S_{i,\lambda_i} = \frac{\exp\{\sum_{j \in N_i} \sum_{\lambda_j \in \Omega} P(\theta_j = \lambda_j) R(\lambda_i, \lambda_j) W_{ij}\}}{\sum_{\lambda_i \in \Omega} \exp\{\sum_{k \in N_i} \sum_{\lambda_k \in \Omega} P(\theta_k = \lambda_k) R(\lambda_i, \lambda_k) W_{ik}\}} \quad (1)$$

where  $R(\lambda_i, \lambda_j)$  are the elements of the label compatibility matrix  $R$ . Here the proximity weights  $W_{ij}$  using a Gaussian function, and are used to weight the label-support. The label probabilities are then iteratively updated using the formula:

$$P^{(n+1)}(\theta_i = \lambda) = \frac{P^{(n)}(\theta_i = \lambda) + \mu S_{i,\lambda}^{(n)}}{\sum_{\lambda_i \in \Omega} (P^{(n)}(\theta_i = \lambda) + \mu S_{i,\lambda}^{(n)})} \quad (2)$$

where  $\mu$  is a constant parameter and  $n$  is the iteration index.

### 3 Kernel Spectral Matching

The problem of point pattern matching is that of establishing one-to-one point correspondences between the two data-sets  $\mathbf{y}$  and  $\mathbf{x}$  extracted from two different images. Ideally, outliers (i.e. extraneous points due to noise) can be removed from the data-sets during matching. Graph spectral methods solve the point correspondence problem by first constructing a weighted graph representation  $G(V, W)$  for each data-set, where  $V$  is the node set (the image points) and  $W$  is the weighted proximity matrix for the nodes that captures the pairwise spatial relationships between image points. One way of constructing the proximity matrix is to use adjacency relationships. Accordingly  $W_{ij} = 1$  if the two points are connected by an edge in the graph, and  $W_{ij} = 0$  otherwise. Another popular choice is to use the Gaussian function  $W_{ij} = e^{-d_{ij}^2/\sigma}$  (e.g. [14] and [13]). Here  $d_{ij}$  is the

Euclidean distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , and  $\sigma$  is a constant. In [14], this is explained as to mapping the original 2-D data to a higher dimensional space to capture the structural arrangement of the feature points. Feature correspondences are then found by using eigendecompositions of the proximity weight matrices. When viewed from the perspective of kernel PCA [12], applying a dissimilarity or similarity function to the original data set is equivalent to the process of using a kernel function to map the data into a higher, possibly infinite, dimensional space. Ideally, this mapping interpolates the data in the new space in a manner that is transformationally invariant. Kernel PCA thus appears to provide us with a theoretical basis for spectral pattern matching. Our aim in this paper is to construct a kernel matrix representation that is further constrained by label consistency information. Our idea is to take advantage of kernel PCA to deliver a more stable and more efficient matching process for articulated point matching.

Kernel PCA [12] can be regarded as a non-linear generalization of the conventional linear PCA method. Conventional PCA provides an orthogonal transformation of the data from a high dimensional space to a low dimensional one, which maximally preserves the variance of the original data. This is done by extracting the first few leading eigenvectors from the data-set covariance matrix, and projecting the data onto these eigenvectors. By contrast, kernel PCA first uses a mapping  $\mathcal{T} : \mathbf{x} \mapsto \Phi(\mathbf{x})$  of the data from the original space into a new feature space  $\mathcal{F}$  of higher, possibly infinite, dimension before extracting the principal components. In practice, an explicit mapping  $\mathcal{T}$  does not always exist so the mapping is performed implicitly by choosing a suitable kernel function  $K(\mathbf{x}_i, \mathbf{x}_j)$ . The kernel  $K$  satisfies Mercer’s theorem [16]. To extract the principal components of the mapped data, first a covariance matrix needs to be constructed for the data in the feature space  $\mathcal{F}$ . Suppose that the image data in the space  $\mathcal{F}$  is centred, then the corresponding covariance matrix is:  $\bar{\mathbf{C}} = \frac{1}{m-1} \sum_{i=1}^m \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^T$ . In [12] Schölkopf, Smola, and Müller show that by solving the eigen-equation  $m\lambda\alpha = K\lambda$ , the  $p_{th}$  feature vector, corresponding to the projection of the  $p_{th}$  feature point on the eigenspace, takes the form

$$\langle v^p, \Phi(\mathbf{x}) \rangle = \frac{1}{\sqrt{\lambda^p}} \sum_{i=1}^m \alpha_i^p k(\mathbf{x}_i, \mathbf{x}) = \sqrt{\lambda^p} \alpha_n^p. \tag{3}$$

To generalize the method to non-centered data, the kernel function  $K$  becomes [12]  $K' = (I - ee^T)K(I - ee^T)$  where  $e = M^{-1/2}(1, 1, \dots, 1)^T$ . When more than one rigid component is present in the data, each component must be centered onto its own respective subpart centre of movement. To do this, we first compute the mean position (i.e. subgroup centre) corresponding to each label. For the group with label  $\lambda$ , the mean position is given by  $\mu_\lambda = \frac{1}{\sum_i P(\theta_i=\lambda)} \sum_i \Phi(\mathbf{x}_i)P(\theta_i = \lambda)$ . The covariance matrix then becomes  $\bar{\mathbf{C}}_{new} = \frac{1}{m-1} \sum_{i=1}^m \tilde{\Phi}(\mathbf{x}_i)\tilde{\Phi}(\mathbf{x}_i)^T$ , and:

$$\begin{aligned} \tilde{\Phi}(\mathbf{x}_i) \cdot \tilde{\Phi}(\mathbf{x}_i)^T &= (\Phi(\mathbf{x}_i) - \sum_\lambda \mu_\lambda P(\theta_i = \lambda))(\Phi(\mathbf{x}_i) - \sum_\lambda \mu_\lambda P(\theta_i = \lambda))^T \\ &= K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{\lambda \in \Omega} \frac{P(\theta_i=\lambda)}{\sum_j P(\theta_j=\lambda)} \sum_j P(\theta_j = \lambda)K(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - \sum_{\lambda \in \Omega} \frac{P(\theta_i=\lambda)}{\sum_k P(\theta_k=\lambda)} \sum_k P(\theta_k = \lambda)K(\mathbf{x}_k, \mathbf{x}_i) \\ &\quad + \sum_{\lambda \in \Omega} \frac{P^2(\theta_i=\lambda)}{\sum_j P(\theta_j=\lambda)\sum_k P(\theta_k=\lambda)} \sum_{j,k} P(\theta_j = \lambda)P(\theta_k = \lambda)K(\mathbf{x}_k, \mathbf{x}_j) \end{aligned} \tag{4}$$

To perform articulated feature point matching, the Gram matrix can be further re-organized to cope with the relative motion of the components. In particular, we weight the kernel matrix  $\tilde{K}$  using label consistency information for each point-set using the formula:

$$\tilde{K}_{ij} = \sum_{l=1}^L P(\theta_i = l)P(\theta_j = l)K_{ij} \quad (5)$$

The covariance matrix  $\overline{C}_{new}$  is then computed using (4) with the above resulting matrix in place of the original matrix  $K$  and its eigen-decomposition computed. The mapping of the feature vectors  $\tilde{\mathbf{y}}_j$  and  $\tilde{\mathbf{x}}_i$  are thus computed by using (3) for the respective model and data point-sets. The next step is to compute an association matrix to measure the similarity of each point pair. Assuming the labels on each feature point are independent of each other, the association of the two feature vectors  $\mathbf{y}_i$  and  $\mathbf{x}_j$  is computed as follows:

$$M_{ij} = \sum_{\lambda=1}^L P(\theta_i = \lambda)P(\theta_j = \lambda) \exp\{-d_{ij}^2/\sigma\}. \quad (6)$$

The correspondences are defined as the most similar node pairs. That is, the correspondence for each node  $\mathbf{x}_i$  in  $\mathbf{y}$  is the node  $\mathbf{y}_j = \max_j M_{ij}$ .

The matching process is an iterative one in which at each step new label probabilities are incorporated to improve matching. As an increasing number of correspondences are found, the value of the quantity  $S = \sum_i e^{-d_{\mathbf{x}_i, \mathbf{y}_j}^2}$ , will increase and approach a maximum value. We thus use  $S$  as a stopping criterion for the iteration process. The matching process is summarised as follows:

1. Initialize  $P$ , threshold =  $t$ ;
2. if  $L > 1$  learn the label semantics from  $\mathbf{y}$ ;
3. Compute the Gaussian association matrix  $W$  for  $\mathbf{x}$ ;
4. Run the labeling process, compute  $P^{new}$ ;
5. Use  $P^{new}$  to compute  $\overline{C}_{new}$  using (4) and  $M$  using (6);
6. Compute  $\mathbf{y}_j = \max_k M_{ik}$  for each point  $\mathbf{x}_i \in \mathbf{x}$ ;
7. diff =  $S - S_{old}$ ;  
return if (diff <  $t$  or iteration > limit); else update  $P$ ;
8. Go to step 3.

## 4 Experimental Results

Experiments are performed with both synthetic and real world data. In the rigid motion case, we also compare the proposed algorithms with the algorithms in [14], [13], and the classical MDS [6]. In both rigid and articulated motions, the experiments focus on the performance of the algorithms when the data are subjected to transformations and uncertainties.

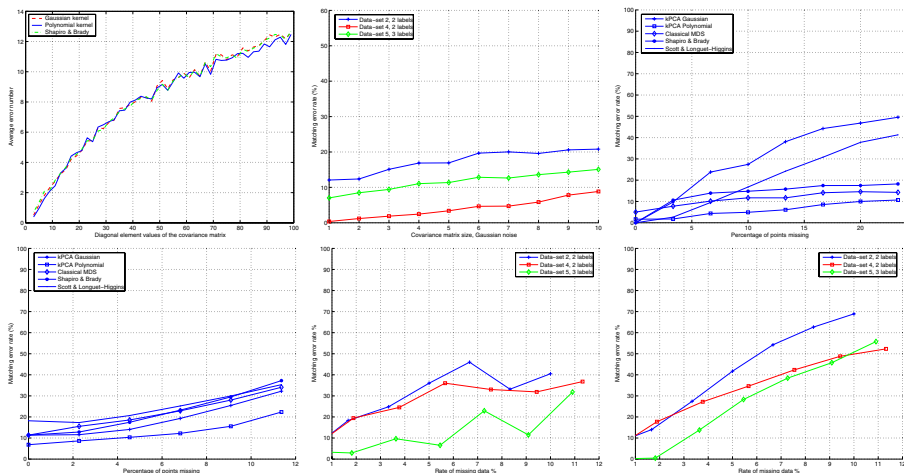
We commence by experimenting on synthetic data. Assume that the point sets are subject to a 2-D affine transformation. Given a point-set  $X = \{\mathbf{x}_i\}_{i=1}^n$  sampled from a rigid object, a synthetic dataset  $X' = sRX + \mathbf{t}$  is generated with a predefined parameter set  $\Theta = (s, \theta, t_x, t_y)$  with  $s$  a scalar,  $(t_x, t_y)$  the translation vector, and  $\theta$  the rotation angle in  $R$ . For single components, the second and third rigid components in the multi-label case, the transformation parameter vector  $\Theta$  is set to  $(0.8, 20^\circ, 10, 15)$ ,  $(0.8, 20^\circ, 10, 15)$ , and  $(1.2, 30^\circ, 10, 15)$ , respectively. For the synthesized single component data-set pair, all algorithms give a 100% correct matching except MDS (with 5% error rate). For articulated motion, the matching process involves a label process to label each feature point  $\mathbf{x}'_i$  to the corresponding rigid component it comes from. The initial label probabilities are assigned uniformly and the results are listed in Table 4. The experiments on feature points with Gaussian random position jitter is evaluated by first adding a randomly generated 2-D Gaussian distributed noise matrix  $D \sim N(\mu, \Sigma)$  on the data point-set; that is,  $\mathbf{x}' = \mathbf{x} + D$ . Results are displayed in Figure 1. The experimental results are the averages of 100 runs for each different  $\Sigma$ . In the left of the figure, we compare the results of using different matching methods on single component point-sets. The right of the figure is for articulated case with synthetic data-sets and also real data-sets.

To simulate structural errors firstly we delete  $l$  consecutive points to simulate occlusion, and secondly we delete  $l$  points in random locations to simulate the effect of segmentation errors for experiments on both synthetic and real data-sets. The matching results are shown in Figure 1. Experiments on real world data-sets include first a sequence of infra-red images of a hand (see, in [2]) which has small geometric deformations and a sequence of the CMU house ([2]) in which the point-sets are of different size and has significant positional jitter. Secondly, a pair of images with two rectangular objects moving away from each other, and images of spectacles with moving arm (data-set 4 and 5, Tabel 4, respectively) are included in the experiments. These results are shown in Tables 1 and 4.

From these experiments, it is clear that the kernel PCA approach gives good results when compared with the approaches of Shapiro & Brady [14], Scott & Longuet-Higgins [13], and the MDS method. Moreover, the kernel method is less sensitive to uncertainties than the alternatives.

**Table 1.** Matching results (Single component, *Numbers of errors*)

Frames	Hand data				CMU House				
	08/25	09/11	09/25	11/25	01/02	01/03	01/04	01/05	01/06
KPCA,Gaussian	6	4	4	11	2	4	2	2	7
KPCA,Polynomial	5	7	6	12	4	5	3	5	13
MDS	35	5	26	27	5	5	25	25	28
Shapiro&Brady	9	6	8	17	3	5	2	2	9
SLH	4	3	5	10	7	6	3	7	9



**Fig. 1.** From top to bottom, left to right: Gaussian jitter, single label synthetic data; Gaussian jitter, multi-label; occlusion, single label synthetic data; occlusion, single label hand sequence; random point missing, multi-label; occlusion, multi-label

**Table 2.** Matching and labeling results II (Gaussian kernel, *error%*)

Data-set	Num of points	Num of labels	No Label Information	Articulated matching(1)*	Articulated matching(2)**	Labeling
1	10	3	10	0	0	0
2	60	2	53.33	13.33	13.33	0
3	31	2	35.48	0	0	0
4	55	2	18.18	7.27	3.64	1.82
5	53	3	45.28	81.13	3.77	7.55

Note: \*: Results based on label information obtained from the label process;  
 \*\*: Results obtained when correct label information is assumed.

## 5 Conclusions

In this paper we have made two contributions. First, we show how the point proximity matrix can be incorporated into the definition of the support function for relaxation labelling and how the label probabilities can be updated with the resulting support. Our second contribution has been to show how label compatibility coefficients can be used to refine the computation of the kernelised proximity matrix for the problems of rigid and articulated point pattern matching. Experimental results reveal that the method offers performance advantages over a number of alternative methods. and gives useful results when there are different moving components in a scene. In the rigid motion case the performance of our algorithm is also comparable to the iterative approaches described in [1,2].

The kernel function used in this paper are possibly not the best choice. One of our future plans to explore kernels that are more stable to structural errors. Our second plan is to do more work on the label process and its interaction with matching. One possibility which has a natural assonance with the kernel method, is to use the heat equation and its spectral solution to model the evolution of label probabilities with time. Work aimed at investigating these points is in-hand and will be reported in due course.

## References

1. M. Carcassoni and E. R. Hancock. "Spectral correspondence for point pattern matching". *Pattern Recognition*. 36(2003) pp.193-204
2. M. Carcassoni and E. R. Hancock. "Correspondence matching with modal clusters". *IEEE Tran. PAMI* vol.25 no.12, 2003
3. H. Chui and A. Rangarajan. "A new point matching algorithm for non-rigid registration", *Computer Vision and Image Understanding*, 89:114-141, 2003.
4. Fan R. K. Chung. "Spectral Graph Theory". *Amer. Math. Soc.*, 92, 1997.
5. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Training Models of Shape from Sets of Examples". *In Proceedings BMVC*, pp.9-18, 1992.
6. T.F.Cox and M.A.A.Cox. "Multidimensional Scaling". *Chapman and Hall*, 1994.
7. J. Kittler and E. R. Hancock. "Combining Evidence In Probabilistic Relaxation". *Intern. Jour. Patt. Recog. And Arti. Intel.*, vol.3, no.1, pp.29 - 51, 1989.
8. S. Pappu, S. Gold and A. Rangarajan. "A framework for non-rigid matching and correspondence". *Advances in Neural Information Processing Systems 8*, 1996.
9. M. Pelillo and M. Refice. "Learning Compatibility Coefficients for Relaxation Labeling Processes". *IEEE Trans. PAMI*, vol.16, no.9, pp.933 - 945, 1994.
10. M. Pilu. "A direct method for stereo correspondence based on singular value decomposition". *IEEE CVPR*, pp.261-266, 1997.
11. A. Rosenfeld and R. Hummel and S. Zucker. "Scene labeling by relaxation operations". *IEEE Trans. Systems. Man and Cybernetics*, vol.6, pp.420 - 433, 1976.
12. B. Schölkopf, A. J. Smola and K. R. Müller. "Nonlinear component analysis as a kernel eigenvalue problem". *Neural Computation*, vol.10, pp.1299-1319, 1998.
13. G. L. Scott and H. C. Longuet-Higgins. "An Algorithm for Associating the Features of Two Images". *Proc. Royal Soc. London Series B*, vol.244, pp.21 - 26, 1991.
14. L. S. Shapiro and J. M. Brady. "Feature-Based Correspondence - An Eigenvector Approach". *Image and Vision Computing*, vol.10, pp.283-288, 1992
15. C. Tomasi and T. Kanade. "Shape and motion from image streams under orthography - A factorization method". *Tech. Rept TR-92-1270, Cornell University*, 1992.
16. V. N. Vapnik. "Statistical learning theory". *John Wiley & Sons, Inc.*, 1998