

On the Use of the Chi-squared Distance for the Structured Learning of Graph Embeddings

Haifeng Zhao*, Antonio Robles-Kelly^{†‡§} and Jun Zhou^{†‡§}

*School of Computer Science and Technology,

Nanjing University of Science and Technology, Nanjing 210094, China

[†]Canberra Research Laboratory, NICTA,

Locked Bag 8001, Canberra ACT 2601, Australia

[‡]College of Engineering and Computer Science,

The Australian National University, Canberra, ACT 0200, Australia

[§]School of Engineering and Information Technology,

UNSW@ADFA, Canberra ACT 2600, Australia

Abstract—In this paper, we describe the use of concepts from the areas of structural and statistical pattern recognition for the purposes of recovering a mapping which can be viewed as an operator on the graph attribute-set. This mapping can be used to embed graphs into spaces where tasks such as classification and retrieval can be effected. To do this, we depart from concepts in graph theory so as to introduce mappings as operators over graph spaces. This treatment leads to the recovery of a mapping based upon the graph attributes which is related to the edge-space of the graphs under study. As a result, the recovered mapping is a linear operator over the attribute set which is associated with the graph topology. To recover this mapping, we employ an optimisation approach whose cost function is based upon the Chi-squared distance and is related to the target function used in discrete Markov Random Field approaches. Thus, the method presented here provides a link between concepts in graph theory, statistical inference and linear operators. We illustrate the utility of the recovered embedding for purposes of shape categorisation and retrieval. We also compare our results to those yielded by alternatives.

I. INTRODUCTION

The problem of embedding relational structures onto a manifold for the purposes of representation, classification, retrieval or visualisation is an important one in computer science which arises in parameterization of three-dimensional data [1], multidimensional scaling (MDS) [2], graph drawing [3] and structural graph matching [4]. In the pattern analysis community, there has recently been renewed interest in the use of embedding methods motivated by graph theory. One of the best known of these is ISOMAP [5]. Related algorithms include locally linear embedding which is a variant of PCA that restricts the complexity of the input data using a nearest neighbor graph [6], and the Laplacian eigenmap that constructs an adjacency weight matrix for the data-points and projects the data onto the principal eigenvectors of the associated Laplacian matrix [7]. Collectively, these methods are sometimes referred to as manifold learning theory.

Embedding methods can also be used to transform the relational-matching problem into one of point-pattern matching in a high-dimensional space. The problem is to find matches between pairs of point sets when there is noise,

geometric distortion and structural corruption. The problem arises in shape analysis [8], motion analysis [9] and stereo reconstruction [10]. As a result, there is a considerable literature on the problem, and many contrasting approaches, including search [11] and optimisation [12] have been attempted. However, the main challenge in graph matching is how to deal with differences in node and edge structure. One of the most elegant recent approaches to the graph matching problem has been to use graph spectral methods [13], and exploit information conveyed by the eigenvalues and eigenvectors of the adjacency matrix. More recently, Sebastian and Kimia [14] have used a distance metric analogous to the string edit distance to perform object recognition from a data-set of shock graphs.

The main argument levelled against the work mentioned above is that it adopts a heuristic approach to the relational matching problem by making use of a goal-directed graph similarity measure. To overcome this problem, several authors have adopted a more general approach using ideas from information and probability theory. For instance, Wong and You [15] defined an entropic graph-distance for structural graph matching. Christmas, Kittler and Petrou [16] have shown how a relaxation labeling approach can be employed to perform matching using pairwise attributes whose distribution is modeled by a Gaussian. Wilson and Hancock [17] have used a MAP (maximum *a posteriori*) estimation framework to accomplish purely structural graph matching. Recently, Caetano *et al.* [18] have proposed a method to estimate the compatibility functions for purposes of learning graph matching.

In this paper, we aim at estimating a linear mapping so as to embed a graph into a high-dimensional space where distances between nodes correspond the structural differences between graphs. This can be viewed as an statistical learning process in which the aim of computation is the recovery of a linear operator which maps the attribute-set of a graph onto a space so as to minimise a cost function arising from a Markovian formulation. In this manner, the recovered mapping is related to the space defined by the graph edge-set while being an operator on the attribute-set. This is important since such

an embedding permits the use of metrics in the target space for purposes of retrieval and categorisation tasks. Moreover, it provides a link between structural and statistical pattern recognition techniques through the use of algebraic graph theory [19], graph spectral methods [13] and Markov Random Fields [20].

II. GRAPH THEORY AND STATISTICAL LEARNING

As mentioned earlier, we aim at learning a linear mapping $\mathcal{T} : \mathcal{A}_i \mapsto \mathfrak{R}^\Omega$ that can be used to embed the attributes of the graph-vertices into a space of dimensionality Ω whose basis is the optimal transformation of a linear map from the vertex to the edge space. In this manner, the embedding will reflect the structure of the edge-space of the graph while being based upon its attribute-set. This has two main advantages. Firstly, the target space for the learned mapping will reflect the structure of the graphs under study. Since similar graphs should have akin edge-spaces, this provides an embedding that is inherently related to a graph topology common to the set Γ . Secondly, note that the mapping \mathcal{T} embeds the vertex-attributes into the graph edge-space according to a linear operator drawn from spectral geometry. This is not only practically useful but theoretically important since it provides a link between the spectra of graphs and linear operators.

A. The Attribute Mapping

Here we work with a data set Γ of attributed graphs. Recall that the aim is to employ structured learning to recover the linear mapping \mathcal{T} so as to embed the attributes for the nodes in the graphs under study into a space whose basis corresponds to the edge-space of the graphs in Γ . Let $G_i = (\mathcal{V}_i, \mathcal{E}_i, \mathcal{A}_i)$ denote the graph indexed i in Γ with node-set $\mathcal{V}_i = \{V_{i,1}, \dots, V_{i,|\mathcal{V}_i|}\}$, edge-set $\mathcal{E}_i = \{\mathcal{E}_{a,b} | V_{i,a}, V_{i,b} \in \mathcal{V}_i\}$ and attribute-set $\mathcal{A}_i = \{A_{i,1}, \dots, A_{i,|\mathcal{V}_i|}\}$.

To commence, we require some formalism. Let the weight matrix \mathcal{W} with elements $\mathcal{W}(a, b)$ for the graph G_i be related to the un-normalised Laplacian $\mathcal{L} = \mathbf{D} - \mathcal{W}$, where \mathbf{D} is a diagonal matrix such that $\mathbf{D} = \text{diag}(\text{deg}(1), \text{deg}(2), \dots, \text{deg}(|\mathcal{V}_i|))$ and $\text{deg}(b) = \sum_{a=1}^{|\mathcal{V}_i|} \mathcal{W}(b, a)$ is the degree of the node indexed b in the graph [13]. Consider the mapping \mathcal{I} of all functions $f(\cdot)$ over the set of vertices \mathcal{V}_i to the functions $g(e)$ over the set of edges \mathcal{E}_i . The incidence mapping \mathcal{I} is then an operator such that $\mathcal{I}g(e) = f(e_+) - f(e_-)$, where the nodes $v = e_+$ and $w = e_-$ are the head and tail, respectively, of the edge $e \in \mathcal{E}_i$. As a result, \mathcal{I} is a $|\mathcal{V}_i| \times |\mathcal{E}_i|$ matrix which satisfies

$$\mathcal{L} = \mathcal{I}\mathcal{I}^T \quad (1)$$

Note that the incidence mapping \mathcal{I} is independent of the orientation of the edges in \mathcal{E}_i . Moreover, it is an operator, so it is independent of the vertex-basis, i.e. it's permutation invariant [19]. With the incidence mapping above, we can now introduce the attribute mapping $Q : \mathcal{A}_i \mapsto \mathcal{E}_i$ such that $S = Q \circ \mathcal{I}^*$, where $S : \mathcal{V}_i \mapsto \mathcal{A}_i$ is the one-to-one mapping between the node-set and the attribute-set of G_i and \mathcal{I}^* is the adjoint of the incidence mapping. In other words, the attribute mapping Q is a linear operator that maps functions over the set of

attributes \mathcal{A}_i to the functions $g(e)$ over the set of edges \mathcal{E}_i . The adjoint of the incidence mapping \mathcal{I}^* maps the functions $g(e)$ to the space of functions $f(\cdot)$ over the vertex-set \mathcal{V}_i .

B. Structured Learning

Note that, so far, we have focused in the attribute and incidence mappings. For a space of vectors, we can view these mappings as matrices which determine an embedding through an algebraic operation. Moreover, note that every incidence and attribute mappings are graph-dependent. This is, they are matrices whose dimensionality is given in terms of the quantities $|\mathcal{V}_i|$ and $|\mathcal{E}_i|$. Here, we aim at learning a global mapping \mathcal{T} which is a matrix whose dimensionality is $|\mathcal{A}_{i,a}| \times \Omega$. In other words, we aim at recovering an operator which can embed any of the attributes for a given $G_i \in \Gamma$ into a space \mathfrak{R}^Ω .

Viewed in this manner, the aim of computation is the recovery of the optimal transformation matrix over the field of attributes for the graphs in Γ . To recover this transformation matrix, we provide a link to Markov Random Field (MRF) models so as to abstract the problem into a graphical setting. To commence, we associate each $V_{i,a} \in \mathcal{V}_i$ with a hidden variable X_a in the state space Λ . The joint probability distribution represented by the MRF is given by

$$P(\mathcal{X}) = \frac{1}{Z} \prod_{G_i, G_k \in \Gamma} \varphi_{i,k}(X_a, X_b) \prod_{G_i \in \Gamma} \zeta_i(X_a) \quad (2)$$

where $\mathcal{X} = \{X_a\}_{a=1, \dots, |\mathcal{V}_i|}$ is the set of hidden variables and $\zeta_i(X_a)$ and $\varphi_{i,k}(X_a, X_b)$ are unitary and binary potential functions which determine the likelihood of the graphs in the data set corresponding to the state $\varrho \in \Lambda$. In the equation above, $Z = \int_{\mathcal{X}} \prod_{G_i, G_k \in \Gamma} \varphi_{i,k}(X_a, X_b) \prod_{G_i \in \Gamma} \zeta_i(X_a)$ is the normalisation factor.

Since this normalisation factor is invariant with respect to X_a , the inference of the above MRF model can be recast as an Maximum A Posteriori (MAP) estimation problem to maximise the joint probability $P(\mathcal{X})$ over the state space Λ . Moreover, we can view the variables X_a in the MRF as continuous vector variables $X_a = [x_{i,1}, \dots, x_{i,|\mathcal{V}_i|}]^T$, where we can view the elements of X_a as the linear product such that $x_{i,a} = \mathcal{T}A_{i,a}$. In other words, the hidden variables correspond to the embeddings of the graph-vertex attributes onto the space defined by the linear mapping $\mathcal{T} : \mathcal{A} \mapsto \mathfrak{R}^\Omega$. From an alternative viewpoint, we can consider $x_{i,a}$ to be the weighted analogue of the attribute-vector for the a^{th} vertex in the graph indexed i in Γ .

Taking the logarithm of Equation 2, we have

$$\begin{aligned} \log P(\mathcal{X}) &= \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \langle c_i(a), x_{i,a} \rangle \\ &+ \sum_{i=1}^{|\Gamma|} \sum_{k=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} w_{i,k}(a, b) \langle x_{i,a}, x_{k,b} \rangle \end{aligned} \quad (3)$$

where $c_i(a) = \log \zeta_i(a)$ and $w_{i,k}(a, b) = \log \varphi_{i,k}(a, b)$ are determined by the potential functions and $\langle \cdot, \cdot \rangle$ denotes the dot product.

Note that, following the previous section, we can define the unary potential $\zeta_i(x_a)$ as an exponential family over the optimal linear transformation of the incidence mapping \mathcal{I}_i into a space in \mathbb{R}^Ω . Thus, in practice, we can recover the potential $\zeta_i(a)$ making use of the Principal Component Analysis (PCA) [2] of the matrix \mathcal{I}_i . Our choice of unary potential hinges in the developments in [21]. It can be shown that the space spanned by the PCA analysis is equivalent to the vertex-to-node scatter for the graph. Thus, we can view the terms $\langle c_i(a), x_{i,a} \rangle$ as the projections of the vectors $x_{i,a}$ onto the subspace defined by the principal directions of the covariance for the mapping between the sets \mathcal{V}_i and \mathcal{E}_i in G_i .

Maximising the above cost function is equivalent to solving the original MRF inference problem, as defined in Equation 2. The cost function is in quadratic form and, hence, it is a natural choice to apply quadratic programming techniques to solve the relaxation problem. However, the Hessian of Equation 3 is determined by the coefficients of the second order term $w_{i,k}(a,b)$ and, as a result, are not necessarily convex. A number of techniques have been proposed to relax the discrete problem above and convert the MRF cost functional into more tractable forms. Along these lines, some examples of this are SDP [22]–[24], SOCP [25], and spectral relaxation [26].

Instead of finding a continuous relaxation for the original cost function of the MRF model, we propose an alternative cost function which is closely related to it. Notice that the first and the second terms on the right-hand-side of the cost function in Equation 3 can be treated as correlation terms. The first of them measures the correlation between the graph and the single node potential. The second term measures the compatibility between graphs and the pairwise node-potential $w_{i,k}(a,b)$. By thinking of correlation as a measure of similarity and viewing as an inverse distance, we can transform the maximisation problem at hand into a minimisation one. To do this, the L_2 norm is a natural choice. The corresponding cost function is hence defined as follows

$$\begin{aligned} \min f(x) = & \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \|c_{i,a} - x_{i,a}\|^2 \\ & + \eta \sum_{i=1}^{|\Gamma|} \sum_{i=k}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} \|w_{i,k}(a,b)\|^2 \|x_{i,a} - x_{k,b}\|^2 \end{aligned} \quad (4)$$

where η is a regularisation constant, for the sake of consistency, we have used vector norms where appropriate.

The reformulation of the cost function as above has several appealing properties. First, it is closely related to the MRF model in terms of its physical meaning. Like the MRF, our cost function also accommodates two complementary terms, i.e. a term which measures the compatibility between the data and its transformed field variable and a smoothness term which can be used to enforce the consistency between the variables for those nodes corresponding to the same graph, i.e. $i = k$. The main difference in the cost functions above is the replacement of the inner product with squared distance. More

importantly, the cost function defined above is convex. This leads to two desirable consequences. Firstly, we can always attain globally optimal solutions for the relaxed problem on the continuous variables. Secondly, the problem can be reduced to that of solving a sparse linear system of equations with positive semidefinite Hessian.

III. THE χ^2 DISTANCE

In this section, we explore the use of the χ^2 distance for purposes of recovering the mapping \mathcal{T} . The reasons for choosing the χ^2 distance hinges in the widespread use of the χ^2 goodness of fit test (also known as Pearson's test [27]) as one of the most common for statistical conformity. The chi-squared distance has found applications in shape recognition [28]. Making use of the chi-squared distance as defined in [28] the cost function then becomes

$$\begin{aligned} \min f(x) = & \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{l=1}^{|\mathcal{C}_{i,a}|} \frac{(c_{i,a}(l) - x_{i,a}(l))^2}{c_{i,a}(l) + x_{i,a}(l)} \\ & + \eta \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} \sum_{l=1}^{|\mathcal{C}_{i,a}|} \frac{(x_{i,a}(l) - x_{k,b}(l))^2}{x_{i,a}(l) + x_{k,b}(l)} \end{aligned} \quad (5)$$

where, as before, we have set the weight $w_{i,k}$ to unity if and only if two nodes indexed a and b correspond to the graph G_i . Also, note that, in the equation above, we have substituted the vector norms for the squares so as to compute the χ^2 distance on a per-element basis and introduced the sum over the index l accordingly.

The χ^2 distance is non-Euclidean and, hence, the cost function above does have a non-linear behaviour in \mathcal{T} . As a result, it cannot be solved in an algebraic fashion as effected in the case of the L_2 -norm in the section above. Nonetheless, we can use, as an alternative to Equation 5, the approximation

$$\begin{aligned} \min f(x) = & \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{l=1}^{|\mathcal{C}_{i,a}|} \frac{(c_{i,a}(l) - x_{i,a}(l))^2}{c_{i,a}(l) + \hat{x}_{i,a}(l)} \\ & + \eta \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} \sum_{l=1}^{|\mathcal{C}_{i,a}|} \frac{(x_{i,a}(l) - x_{k,b}(l))^2}{\hat{x}_{i,a}(l) + \hat{x}_{k,b}(l)} \end{aligned} \quad (6)$$

where $\hat{x}_{i,a} = \mathcal{T}^* A(i,a)$ and $\mathcal{T}^* \approx \mathcal{T}$.

This suggests an iterative scheme akin to re-weighted least squares techniques. Thus, we can recover the associated attribute mapping by indexing to iteration number t and using the following update equation

$$\operatorname{argmin}_{\mathcal{T}^{t+1}} f(x) = \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \|\hat{c}_{i,a} - \mathcal{T}^{t+1} \hat{A}_{i,a}\|^2 + \eta \hat{\gamma} \|\mathcal{T}^{t+1}\|^2 \quad (7)$$

In the equation above, the term

$$\eta \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} \sum_{l=1}^{|\mathcal{C}_{i,a}|} \frac{(x_{i,a}(l) - x_{k,b}(l))^2}{\hat{x}_{i,a}(l) + \hat{x}_{k,b}(l)}$$

has been rewritten as

$$\eta \hat{\gamma} \|\mathcal{T}^{t+1}\|^2 = \eta \|\mathcal{T}^{t+1}\|^2 \sum_{i=1}^{|\Gamma|} \sum_{a=1}^{|\mathcal{X}_i|} \sum_{b=1}^{|\mathcal{X}_k|} \|\hat{A}_{i,a}(l) - \hat{A}_{k,b}(l)\|^2$$

where $\hat{A}_{i,a} = A_{i,a} \cdot / (c_{i,a} + \mathcal{T}^t A_{i,a})$, $\hat{c}_{i,a} = c_{i,a} \cdot / (c_{i,a} + \mathcal{T}^t A_{i,a})$ and $\cdot /$ denotes element-wise division of the vectors under consideration. Thus, we can minimise the cost function in Equation 7 making use of an iterative scheme by re-weighting the attributes $\hat{A}_{i,a}$ and variables $\hat{c}_{i,a}$ as an alternative to the straight application of $A_{i,a}$ and $c_{i,a}$, respectively.

IV. IMPLEMENTATION ISSUES

With the ingredients above, we now turn our attention to the implementation and application of our embedding method. Here, training can be viewed as the stage where the linear mapping \mathcal{T} is learned from the graph vertex-attributes and the PCA of the incidence mapping. The testing stage then becomes the use of the mapped graph attributes into the target space for different purposes, for example, categorisation and retrieval.

The training stage starts from constructing the attributes of the graph-vertices. The attribute-set is application dependent. Here, we view, in general, vertex-attributes $A_{i,a}$ as vectors, where each of these has a one-to-one correspondence to a graph vertex. To recover the embedding transformation target $c_{i,a}$, we do the following. Given the graph weight matrix \mathcal{W} , we can compute the Laplacian $\mathcal{L} = \mathbf{D} - \mathcal{W}$. The incidence mapping is recovered via a Young-Householder [29] decomposition on \mathcal{L} . With the incidence mapping \mathcal{I} at hand for every graph in Γ , the quantities $c_{i,a}$ are recovered making use of PCA. This is, we perform an eigendecomposition on \mathcal{I} so as to select the leading N eigenvectors of the incidence mapping. This yields $c_{i,a} = [\phi_1(a), \phi_2(a), \dots, \phi_N(a)]^T$, where $\phi_k = [\phi_k(1), \phi_k(2), \dots, \phi_k(|\mathcal{V}_i|)]^T$ corresponds to the k^{th} eigenvector of the incidence mapping \mathcal{I} for the graph indexed i in Γ . With $c_{i,a}$ at hand, the linear mapping matrix \mathcal{T} can be solved by extremising $f(x)$ as described in the previous sections.

In summary, the step sequence of the method is as follows:

- 1) For every graph in Γ , compute the corresponding incidence mapping \mathcal{I} via the Young-Householder decomposition of the Laplacian \mathcal{L} .
- 2) Compute the quantities $c_{i,a}$ via PCA on the incidence mappings for the graphs in the data set.
- 3) Compute linear mapping \mathcal{T} by extremising the cost functions in Equations 7 as described in Section II-B.

Using the linear mapping matrix \mathcal{T} , we can transform any $A_{i,a}$ into a target space, where each graph is represented as a matrix whose a^{th} row corresponds to the coordinates associated to the attribute indexed a in the i^{th} graph in Γ . As a result, relational matching between graphs can be performed by comparing the distances between the transformed attributes. This is due to the fact that there is a known one-to-one relationship between vertices and attributes in the graph. Further, these row vectors can be used to represent each graph as a probability distribution of pairwise vertex distance in the target space. In practice, these can be done via a histogram of distance frequencies whose bin-centres in the embedding space can be recovered using a clustering method such as k -means or maximum-likelihood estimation (MLE) approaches. This can be viewed as a codebook in the target space. In this

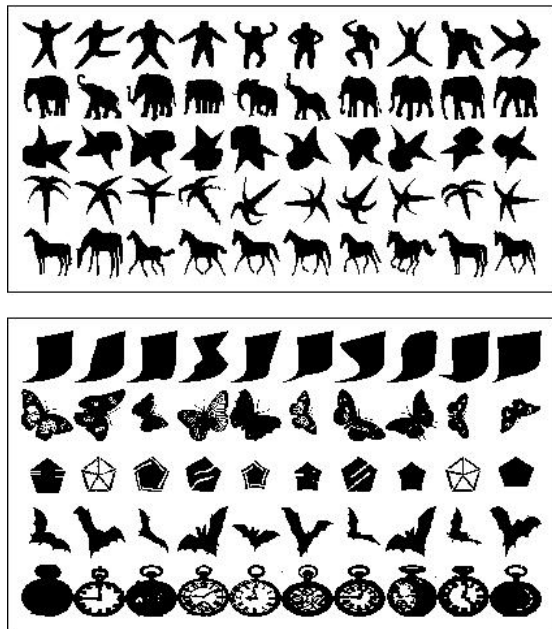


Fig. 1. Samples images from the two shape databases used in our experiments. Top Shape 1000 dataset; Bottom panel: MPEG7 CE-Shape-1 shape dataset

way, we transfer the structural representation of a graph into a statistical representation that can be used for categorisation and retrieval tasks.

V. EXPERIMENTAL RESULTS

Now, we turn our attention to the applicability of the embedding \mathcal{T} to shape retrieval and categorisation. To this end, we have used two datasets. The first of these is the MPEG7 CE-Shape-1 shape database, which contains 1400 binary shapes of 70 different classes with 20 images in each category. The second of these is the collection of 1000 shapes presented in [30]. In Figure 1, we show some example shapes for both databases. We have represented each shape as a graph in Γ whose vertices correspond to contour pixels sampled in regular intervals. Here, we have sampled 1 in every 10 pixels on the shape contours. With the sample contour pixels at hand, we build a fully connected graph whose edge-weights are given by the Euclidean distances on the image plane between each pair of pixel sets. Thus, the entries of the weight matrix for the graph correspond to the pairwise distances between the image-coordinates for every pair of vertices in the graph. The weigh matrix is then normalised to unity so as to have every weight in the graph in the interval $[0, 1]$. The attribute set is given by the frequency histogram of these distances for every clique. That is, for the a^{th} vertex in G_i , the attribute $A_{i,a}$ is given by the histogram for the edge-weights for the clique centered at the node indexed a . In our experiments, we have used 12 bins for the frequency histogram computation.

	MPEG7 CE-Shape-1 database	Data set in [30]
Baseline	76.2%	79.6%
χ^2	80%	85%

TABLE I

GRAPH CATEGORISATION RESULTS ON THE TWO DATASETS. TOP ROW: BASELINE CATEGORISATION RATE. BOTTOM ROWS: CLASSIFICATION RATES USING OUR EMBEDDING FOR THE χ^2 DISTANCE.

A. Categorisation

As mentioned earlier, one of the advantages of the embedding strategy adopted here is that it provides a means to connect structural pattern recognition with statistical pattern recognition. For shape classification, we make use of both datasets. In both cases, we employ the histogram of pairwise distances in the embedding space for the coordinates $x_{i,a} = \mathcal{T}A_{i,a}$ and construct a frequency histogram as a graph feature vector that can be used to obtain a ‘‘codeword’’ for each graph. We have done this so as to compare our results to those yielded by a baseline akin to the bag of words approaches in computer vision [31]. For the baseline, we have used the PCA of the attribute-set, i.e. the matrices A_i so as to recover a codebook via k -means clustering. For our embeddings, we have used the frequency histograms for the distances between pairs of embeddings $x_{i,a} = \mathcal{T}A_{i,a}$ for those attributes in the same graph. These distance histograms have been used to recover a codebook which is, again, computed using k -means clustering. For both, the baseline and our method, we have set $k = 200$. Using the pairwise distances for the histogram representation of graphs, we can construct a codebook for all shapes, we have performed classification making use of a linear SVM. This enables us to perform supervised learning on shapes, and, thus, can take advantage of the recent progresses in machine learning.

For our shape classification experiments, we divided the graphs in the dataset into a training and a testing set. Each of these contains half of the graphs in each dataset. This is, for the MPEG-7 dataset we have used 700 graphs for training and 700 for testing. For the other dataset, we have used 500 graphs for testing and 500 for training. Here we have recovered the embedding matrix \mathcal{T} making use of those graphs in the training set only. For both, the baseline and our method, we have tuned the SVM parameters using ten-fold cross validation. The classification results are shown in Table I. Note that, for our method, the classification rate for the χ^2 distance, represent a margin of advantage with respect to the baseline.

B. Shape Retrieval

Finally, we turn our attention to shape retrieval. For retrieval, we have employed the codewords recovered by the baseline and our embedding used in the previous section and applied a k -nearest neighbour retrieval on the the χ^2 distances between the query and the data-base codewords.

In Figure 2 we show the average precision-recall curves for our method and the embedding recovered by using the cost function in Equation 4, which uses the L2 norm as an

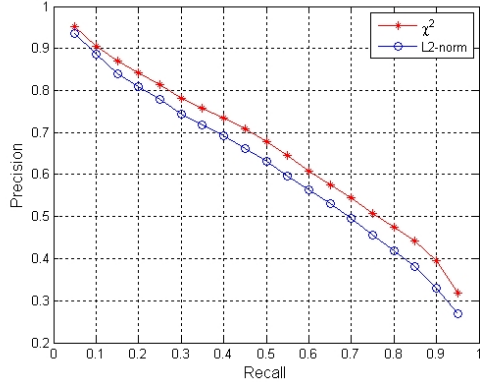


Fig. 2. Precision recall plots for the from the 1000 shape database in [30] for both the χ^2 and the L2 norm.

alternative to the χ^2 distance. For the plots in the figure, we have used the collection of 1000 shapes. It comprises 1000 graphs which we have used for categorisation by randomly selecting one of them as a query, recovering the embedding \mathcal{T} for the rest of the shapes in the database and, via a k -nearest neighbour classifier, recovered the precision for a set recall.

From the plots, we can observe that, for our method, the χ^2 distance provides a margin of advantage over the L2 norm. The retrieval results are comparable to those in Baseski *et al.* [30]. Note that, nonetheless our graphs are built in a very simple manner, we still achieve precision recall rates similar to tree-edit approaches developed specially for binary shapes. Also, note that, the embedding favours the retrieval of graphs whose topology is similar to one another. As the topology becomes more disparate between query and candidate, the precision on the k -nearest neighbour retrieval will decrease.

VI. CONCLUSIONS

In this paper, we have proposed a method to recover a mapping which is based upon the graph attribute-set and, at the same time, is inherently related to the graph topology. We have done this by drawing a link between the incidence mapping and a linear operator over the graph-vertex attributes. This linear operator is, in fact, a mapping that can be used for purposes of embedding graphs in a space where classification and retrieval tasks can be effected. We recover this embedding making use of the Chi-squared distance and statistical learning techniques. Thus, we can view the recovery of the embedding as a learning process over a common topology for the set of graphs under study. Where the learning process is based upon a cost function which has clear links to MRFs and is convex in nature. We exemplify the utility of our method for categorisation and retrieval on two shape databases.

ACKNOWLEDGMENT

This work was done while Haifeng Zhao was a visiting scholar at the Canberra Research Laboratory of NICTA. He is supported by the National Natural Science Foundation of China (NSFC) under grant number 60775015, 60873151 and

90820306, the Priority Academic Program Development of Jiangsu Higher Education Institutions and Qinglan Project. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

REFERENCES

- [1] C. Gotsman, X. Gu, and A. Sheffer, "Fundamentals of spherical parameterization for 3D meshes," *ACM Trans. on Graphics*, vol. 22, no. 3, pp. 358–363, 2003.
- [2] I. Borg and P. Groenen, *Modern Multidimensional Scaling, Theory and Applications*, ser. Springer Series in Statistics. Springer, 1997.
- [3] G. Di Battista, P. Eades, R. Tamassia, and I. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, 1998.
- [4] M. F. Demirci, A. Shokoufandeh, S. Dickinson, Y. Keselman, and L. Bretzner, "Many-to-many feature matching using spherical coding of directed graphs," in *ECCV04*, 2004, pp. I: 322–335.
- [5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [7] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Neural Information Processing Systems*, no. 14, 2002, pp. 634–640.
- [8] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [9] P. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *International Journal of Computer Vision*, vol. 24, pp. 271–300, 1997.
- [10] L. Shapiro and J. M. Brady, "Feature-based correspondance - an eigenvector approach," *Image and Vision Computing*, vol. 10, pp. 283–288, 1992.
- [11] S. Ullman, "Filling in the gaps: the shape of subjective contours and a model for their generation," *Biological Cybernetics*, vol. 25, pp. 1–6, 1976.
- [12] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *PAMI*, vol. 18, no. 4, pp. 377–388, April 1996.
- [13] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.
- [14] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Shock-based indexing into large shape databases," in *European Conference on Computer Vision*, vol. 3, 2002, pp. 731–746.
- [15] A. K. C. Wong and M. You, "Entropy and distance of random graphs with application to structural pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 599–609, 1985.
- [16] W. J. Christmas, J. Kittler, and M. Petrou, "Structural matching in computer vision using probabilistic relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 749–764, 1995.
- [17] R. Wilson and E. R. Hancock, "Structural matching by discrete relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 634–648, June 1997.
- [18] T. Caetano, L. Cheng, Q. Le, and A. Smola, "Learning graph matching," in *Proceedings of the 11th International Conference on Computer Vision*, 2007, pp. 14–21.
- [19] N. L. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 1993.
- [20] P. Bremaud, *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*. Springer, 2001.
- [21] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. of the Int. Conf. on Machine Learning*, 2004, pp. 225–232.
- [22] J. Keuchel, C. Schnorr, C. Schellewald, and D. Cremers, "Binary partitioning, perceptual grouping, and restoration with semidefinite programming," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1364–1379, 2003.
- [23] J. Keuchel, "Multiclass image labeling with semidefinite programming," in *European Conference on Computer Vision*, 2006, pp. 454–467.
- [24] P. Torr, "Solving markov random fields using semi definite programming," in *Intl Workshop on Artificial Intelligence and Statistics*, 2003.
- [25] M. Kumar, P. Torr, and A. Zisserman, "Solving markov random fields using second order cone programming relaxations," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 1045–1052.
- [26] T. Cour and J. Shi, "Solving markov random fields with spectral relaxation," in *Intl. Conf. on Artificial Intelligence and Statistics*, 2007.
- [27] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine*, vol. 50, pp. 157–175, 1900.
- [28] H. Zhang and J. Malik, "Learning a discriminative classifier using shape context distances," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. I:242–247.
- [29] G. Young and A. S. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, pp. 19–22, 1938.
- [30] E. Baseski, A. Erdem, and S. Tari, "Dissimilarity between two skeletal trees in a context," *Pattern Recognition*, vol. 42, no. 3, pp. 370–385, 2009.
- [31] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Comp. Vision and Pattern Recognition*, 2005, pp. II:524–531.