

High Order Structural Matching Using Dominant Cluster Analysis*

Peng Ren, Richard C. Wilson, and Edwin R. Hancock

Department of Computer Science, The University of York, York, YO10 5GH, UK
{pengren, wilson, erh}@cs.york.ac.uk

Abstract. We formulate the problem of high order structural matching by applying *dominant cluster analysis* (DCA) to a direct product hypergraph (DPH). For brevity we refer to the resulting algorithm as DPH-DCA. The DPH-DCA can be considered as an extension of the game theoretic algorithms presented in [8] from clustering to matching, and also as a reduced version of reduced version of the method of ensembles of affinity relations presented in [6]. The starting point for our method is to construct a K -uniform direct product hypergraph for the two sets of higher-order features to be matched. Each vertex in the direct product hypergraph represents a potential correspondence and the weight on each hyperedge represents the agreement between two K -tuples drawn from the two feature sets. Vertices representing correct assignment tend to form a strongly intra-connected cluster, i.e. a dominant cluster. We evaluate the association of each vertex belonging to the dominant cluster by maximizing an objective function which maintains the K -tuple agreements. The potential correspondences with nonzero association weights are more likely to belong to the dominant cluster than the remaining zero-weighted ones. They are thus selected as correct matchings subject to the one-to-one correspondence constraint. Furthermore, we present a route to improving the matching accuracy by invoking prior knowledge. An experimental evaluation shows that our method outperforms the state-of-the-art high order structural matching methods[10][3].

1 Introduction

Many problems in computer vision and machine learning can be posed as that of establishing the consistent correspondences between two sets of features. Traditional matching approaches are usually confined to structures with pairwise relations. Recently, a number of researchers have attempted to extend the matching process to incorporate higher order relations. Zass *et al.* [10] are among the first to investigate this problem by introducing a probabilistic hypergraph matching framework, in which higher order relationships are marginalized to unary order. It has already been pointed out in [1] that this graph approximation is just a low pass representation of the original hypergraph and causes information loss and inaccuracy. On other hand, Duchenne *et al.* [3] have developed the spectral technique for graph matching [4] into a higher order matching

* We acknowledge the financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250). Edwin R. Hancock is supported by a Royal Society Wolfson Research Merit Award.

framework using the so called *tensor power iteration*. Although they adopt an L_1 norm constraint in computation, the original objective function is subject to an L_2 norm and does not satisfy the basic probabilistic properties.

We present a framework based on applying *dominant cluster analysis* (DCA) to a direct product hypergraph (DPH). The idea is to extend the main cluster method of Leordeanu and Hebert [4] for graphs and its generalization for higher order matching [3], using dominant cluster analysis. Furthermore, we present a method for initializing our algorithm that can be used to suppress outliers. This improves the matching performance of our method, and comparable results can not be achieved by using alternative high order matching algorithms [3][10]. Similar ideas have recently been presented in [6]. Our method however, generalises the methods described in [3][10] from graphs to hypergraphs, and is more principled in its formulation.

2 Problem Formulation

We represent the set of K th order feature relationships by a K -uniform hypergraph $HG(V, E)$, whose hyperedges have identical cardinality K . Each vertex $v_i \in V$ in the K -uniform hypergraph $HG(V, E)$ represents one element in the feature set. Each hyperedge $e_i \in E$ represents one K -tuple $\{v_{i_1}, \dots, v_{i_K}\} \in V$ and the weight attached to each hyperedge represents the similarity measure on the K -tuple encompassed by the hyperedge. For simplicity, we denote a vertex v_i by its index i in the remainder of our work. The K -uniform hypergraph $HG(V, E)$ can be represented as a K th order tensor \mathcal{H} , whose element H_{i_1, \dots, i_K} is the hyperedge weight if there is a hyperedge encompassing the vertex subset $\{i_1, \dots, i_K\} \in V$, and zero otherwise. The problem of matching two feature sets both constituted by K th order relationships can then be transformed to that of matching the two associated K -uniform hypergraphs $HG(V, E)$ and $HG'(V', E')$. To this end, we establish the high order compatibility matrix \mathcal{C} , i.e. compatibility tensor, for $HG(V, E)$ and $HG'(V', E')$. The elements of the K th order compatibility tensor \mathcal{C} are defined as follows

$$\mathcal{C}_{i_1 i'_1, \dots, i_K i'_K} = \begin{cases} 0 & \text{if } H_{i_1, \dots, i_K} = 0 \text{ or } H'_{i'_1, \dots, i'_K} = 0; \\ s(H_{i_1, \dots, i_K}, H'_{i'_1, \dots, i'_K}) & \text{otherwise;} \end{cases} \quad (1)$$

where $s(\cdot, \cdot)$ is a function that measures hyperedge similarity. We define the hyperedge similarity using a Gaussian kernel $s(H_{i_1, \dots, i_K}, H'_{i'_1, \dots, i'_K}) = \exp(-\|H_{i_1, \dots, i_K} - H'_{i'_1, \dots, i'_K}\|_2^2 / \sigma_1)$ where σ_1 is a scaling parameter. Many alternative similarity measures can be used instead. Each element of the compatibility tensor \mathcal{C} represents a similarity measure between the two corresponding hyperedges. The hyperedge pair $\{i_1, \dots, i_K\}$ and $\{i'_1, \dots, i'_K\}$ with a large similarity measure has a large probability $\Pr(\{i_1, \dots, i_K\} \leftrightarrow \{i'_1, \dots, i'_K\} | H, H')$ for matching. Here the notation \leftrightarrow denotes a possible matching between a pair of hyperedges or a pair of vertices. Under the conditional independence assumption of the matching process [10], the hyperedge matching probability can be factorized over the associated vertices of the hypergraphs as $\Pr(\{i_1, \dots, i_K\} \leftrightarrow \{i'_1, \dots, i'_K\} | HG, HG') = \prod_{n=1}^K \Pr(i_n \leftrightarrow i'_n | HG, HG')$ where $\Pr(i_n \leftrightarrow i'_n | HG, HG')$ denotes the probability for the possible matching $i_n \leftrightarrow i'_n$

to be correct. For two hypergraphs $HG(V, E)$ and $HG(V', E')$ with $|V| = N$ and $|V'| = N'$ respectively, we denote their $N \times N'$ matching matrix by \mathbf{P} with entries $P_{ii'} = \Pr(i \leftrightarrow i' | HG, HG')$. High order matching problems can be formulated as locating the matching probability that most closely accords with the elements of the compatibility tensor, i.e. seeking the optimal \mathbf{P} by maximizing the objective function

$$\begin{aligned} f(\mathbf{P}) &= \sum_{i_1=1}^N \sum_{i'_1=1}^{N'} \cdots \sum_{i_K=1}^N \sum_{i'_K=1}^{N'} C_{i_1 i'_1, \dots, i_K i'_K} \Pr(\{i_1, \dots, i_K\} \leftrightarrow \{i'_1, \dots, i'_K\} | HG, HG') \\ &= \sum_{i_1=1}^N \sum_{i'_1=1}^{N'} \cdots \sum_{i_K=1}^N \sum_{i'_K=1}^{N'} C_{i_1 i'_1, \dots, i_K i'_K} \prod_{n=1}^K P_{i_n i'_n} \end{aligned} \quad (2)$$

subject to $\forall i, j, P_{ii} \geq 0$ and $\sum_{i=1}^N \sum_{i'=1}^{N'} P_{ii'} = 1$. Let $\widehat{\Pr}(i \leftrightarrow i' | HG, HG') = \widehat{P}_{ii'}$ where $\widehat{P}_{ii'}$ is the (i, i') th entry of $\widehat{\mathbf{P}}$. We refer to $\widehat{\Pr}(i \leftrightarrow i' | HG, HG')$ as the matching probability for vertex i and i' , and the set of matching probabilities $\{\widehat{\Pr}(i \leftrightarrow i' | HG, HG') | i \in V; i' \in V'\}$ obtained by maximizing (2) reveal how likely it is that each correspondence is correct according to structural similarity between the two hypergraphs HG and HG' . This formulation has also been adopted in tensor power iteration for higher order matching [3]. However, the difference between our method and the existing algorithms is that we restrict the solution of (2) to obey the the fundamental axioms of probability, i.e. positiveness and unit total probability mass. This constraint not only provides an alternative probabilistic perspective for hypergraph matching, but also proves convenient for optimization.

Once the set of matching probabilities satisfying (2) are computed, correspondences between vertices drawn from HG and HG' can be established. Matchings with a zero probability are the least likely correspondences, and matchings with nonzero probabilities tend to be those with significant similarity between their structural contexts. Our aim is to seek the subset of possible matchings with nonzero probabilities which satisfy (2) and that are subject to the one-to-one matching constraint.

3 High Order Matching as Dominant Cluster Analysis on a Direct Product Hypergraph

In this section we pose the high order relational matching problem formulated in (2) as one of *dominant cluster analysis* on a *direct product hypergraph*. We commence by establishing a direct product hypergraph for the two hypergraphs to be matched. Optimal matching can be achieved by extracting the dominant cluster of vertices from the direct product hypergraph.

3.1 Direct Product Hypergraph

The construction of a direct product hypergraph for two K -uniform hypergraphs is a generalization of that of the direct product graph [9], which can be used to construct kernels for graph classification. We extend the concept of a direct product graph to

encapsulate high order relations residing in a hypergraph and apply this generalization to hypergraph matching problems. For two K -uniform hypergraphs $HG(V, E)$ and $HG'(V', E')$, the direct product HG_{\times} is a hypergraph with vertex set

$$V_{\times} = \{(i, i') | i \in V, i' \in V'\}; \quad (3)$$

and edge set

$$E_{\times} = \{ \{(i_1, i'_1) \cdots (i_K, i'_K)\} | \{i_1, \dots, i_K\} \in E, \{i'_1, \dots, i'_K\} \in E' \}. \quad (4)$$

The vertex set of the direct product hypergraph HG_{\times} consists of Cartesian pairs of vertices drawn from HG and HG' separately. Thus the cardinality of the vertex set of HG_{\times} is $|V_{\times}| = |V||V'| = NN'$. The direct product hypergraph HG_{\times} is K -uniform, and each K -tuple of vertices in HG_{\times} is encompassed in a hyperedge if and only if the corresponding vertices in HG and HG' are both encompassed by a hyperedge in the relevant hypergraph. Each hyperedge in a direct product hypergraph is weighted by the similarity between the two associated hyperedges from HG and HG' .

Furthermore, from our definition of direct product hypergraph, it is clear that the compatibility tensor \mathcal{C} defined in (1) is in fact the tensor \mathcal{C}_{\times} associated with the direct product hypergraph HG_{\times} for HG and HG' . Every possible matching $i \leftrightarrow i'$ is associated with the vertex (i, i') in HG_{\times} . For simplicity we let α denote a vertex in HG_{\times} instead of (i, i') , and let \mathbb{D} denote the subset of vertices in HG_{\times} which represent the correct vertex matching for HG and HG' . We denote the probability for the vertex α belonging to \mathbb{D} by $\Pr(\alpha \in \mathbb{D} | HG_{\times})$. For a direct product hypergraph with N_{\times} vertices, we establish a $N_{\times} \times 1$ vector \mathbf{p} with its α th element $p_{\alpha} = \Pr(\alpha \in \mathbb{D} | HG_{\times})$. With these ingredients the optimal model satisfying the condition (2) reduces to

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmax}} \sum_{\alpha_1=1}^{N_{\times}} \cdots \sum_{\alpha_K=1}^{N_{\times}} C_{\alpha_1, \dots, \alpha_K} \prod_{n=1}^K p_{\alpha_n} \quad (5)$$

subject to the constraints $\forall \alpha, p_{\alpha} \geq 0$ and $\sum_{\alpha=1}^{N_{\times}} p_{\alpha} = 1$. Following the construction of a direct product hypergraph, the objective function (5) is a natural extension of that in [8] from clustering to matching. It is also a reduced version of the objective function of ensembles of affinity relations [6], with no manual threshold on the optimization.

According to (5), zero probability will be assigned to the vertices that do not belong to \mathbb{D} . We refer to the probability $\hat{\Pr}(\alpha \in \mathbb{D} | HG_{\times}) = \hat{p}_{\alpha}$ where \hat{p}_{α} is the α th element of the vector $\hat{\mathbf{p}}$ satisfying the optimality condition in (5) as the association probability for the vertex α . Therefore, the matching problem can be solved by extracting the cluster of vertices with nonzero association probabilities in the direct product hypergraph.

3.2 Dominant Cluster Analysis

In this subsection, we formulate the problem of high order structural matching by applying *dominant cluster analysis* (DCA) to a direct product hypergraph (DPH). A dominant cluster of a hypergraph is the subset of vertices with the greatest average similarity, i.e. average similarity will decrease subject to any vertex deletion from or vertex addition to

the subset. Drawing on the concept of the dominant set in a graph [7] and its game theoretic generalization [8], we can easily perform DPH-DCA by applying the following update until convergence is reached [2]

$$p_\alpha^{\text{new}} = \frac{p_\alpha \sum_{\alpha_2=1}^{N_\times} \cdots \sum_{\alpha_K=1}^{N_\times} C_{\alpha, \alpha_2, \dots, \alpha_K} \prod_{n=2}^K p_{\alpha_n}}{\sum_{\beta=1}^{N_\times} p_\beta \sum_{\beta_2=1}^{N_\times} \cdots \sum_{\beta_K=1}^{N_\times} C_{\beta, \beta_2, \dots, \beta_K} \prod_{n=2}^K p_{\beta_n}} \quad (6)$$

At convergence the weight \hat{p}_α is equal to the association probability $\hat{\text{Pr}}(\alpha \in \mathbb{D} | HG_\times)$, i.e. the probability for the corresponding potential matching $i \leftrightarrow i'$ to be correct.

4 Matching with Prior Rejections

The high order structural matching algorithm described in Section 3 is a unsupervised process. The weight of each vertex in the direct product hypergraph can be initialized by using a uniform distribution of probability. However, if two vertices in a hypergraph have the same structural context, i.e. their interchange does not change the hypergraph structure, they can cause ambiguity when matching is attempted. Two alternative state-of-the-art methods, namely probabilistic hypergraph matching [10] and tensor power iteration [3], also suffer from this shortcoming.

However, if prior knowledge about outliers (i.e. hypergraph vertices for which no match exists) is available, we can to a certain extent avoid the ambiguity and improve matching accuracy by using a different weight initialization strategy. We refer to the vertex subset $V_\times^o \subseteq V_\times$ (i.e. possible correspondences) associated with available outliers as prior rejections, and the adopted initialization in the light of prior rejections is as follows

$$w(\alpha) = \begin{cases} 0 & \text{if } \alpha \in V_\times^o; \\ 1/(N_\times - N_\times^o) & \text{otherwise;} \end{cases} \quad (7)$$

where N_\times^o is the cardinality of V_\times^o .

The initialization scheme (7) improves the matching accuracy within the DPH-DCA framework because the vertex weight $w(\alpha)$ in the numerator of the update formula (6) plays an important role in maintaining the initial rejection. It enables the prior rejections to maintain a zero weight and does not affect the matching scores for other possible correspondences at each update until converged. The extent to which the matching accuracy can be improved depends on the amount of prior rejections available. The more prior knowledge concerning the outliers that is available, the more accurate the matching that can be obtained. This will be verified in our experimental section.

In [6], the authors have described the same initialization step as a disadvantage. On the other hand, we argue that the initialization scheme (7) does not apply to the alternative methods[10][3] even when identified outliers are available. The probabilistic hypergraph matching method [10] initializes a matching score by a fixed value obtained from the marginalization of the compatibility tensor, and thus can not accommodate the prior rejections by using (7). The tensor power iteration method [3], though manually initialized, converges to a fixed matching score for different initializations.

5 Experiments

We test our algorithm for high order structural matching on two types of data. Firstly, we test our method on synthetic data to evaluate its robustness to noise and outliers. Secondly, we conduct experiments to match features extracted from images. Prior rejections are considered for both types of data to improve the matching accuracy. We compare our method with two state-of-the-art methods, i.e. probabilistic hypergraph matching (PHM) [10] and tensor power iteration (TPI) [3].

5.1 Matching Synthetic Data

We commence with the random generation of a structural prototype with 15 vertices. The distance d_{ij} between each pair of vertices i and j of the prototype is randomly distributed subject to the Gaussian distribution $N(1, 0.5)$. We test our method by establishing correspondences between the prototype structure and a modified structure. The alternative modifications include a) noise addition, b) vertex deletion, c) rescaling and d) rotation. Since neither the probabilistic hypergraph matching method nor the tensor power iteration method relies upon a specific initialization, we test our DPH-DCA matching method without prior rejections to make a fair comparison with these two alternative methods. To test the performance of different methods for hypergraph matching we re-scaled the distance between each of vertex pairs by a random factor and rotate the structure by a random angle. In this case, the pairwise relationships no longer holds for the matching task. We use the sum of polar sines presented in [5] as a high order similarity measure for point tuples. We measure the similarity of every 3-tuple within the vertex set and thus establish a weighted 3-uniform hypergraph for the structure. The compatibility tensor \mathcal{C} for two structures is computed according to (1) with $\sigma_1 = 0.1$. Figure 1(a) illustrates the results of the matching accuracy as a function of noise level. It is clear that our DPH-DCA framework outperforms the two alternative methods at each noise level. To take the investigation one step further, we study the performance of our method for matching structures of different vertex cardinality. To this end, we extract a substructure from a prototype and slightly perturb the distance between each vertex pair by adding random noise normally distributed according to $N(0, 0.04)$. The cardinality of the vertex set of the substructure varies from 14 down to 5. Vertices not in the substructure are outliers for the matching process. For each vertex cardinality of a substructure, 100 trials are performed. Figure 1(b) illustrates the matching accuracy as a function of outlier number for the three methods. It is clear that our DPH-DCA framework outperforms the two alternative methods at each number of outliers. We have also evaluated the matching accuracy of our DPH-DCA framework at different levels of available prior rejection. To this end, we have extracted a 5-vertex substructure from a prototype and slightly perturb the distance between each vertex pair by adding random noise normally distributed according to $N(0, 0.04)$. We involve prior rejections by rejecting the matchings associated with a varying number of outliers. Figure 1(c) illustrates the matching accuracy as a function of the number of rejected outliers. It is clear that the matching accuracy grows monotonically as the number of rejected outliers increases.

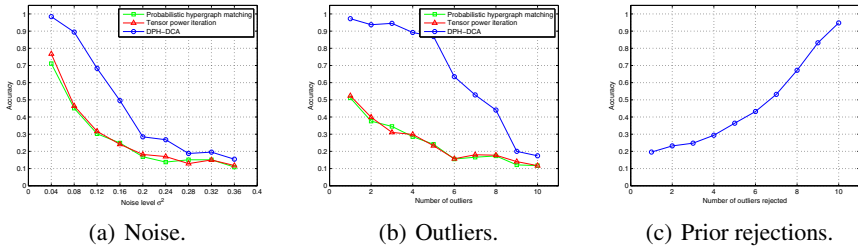
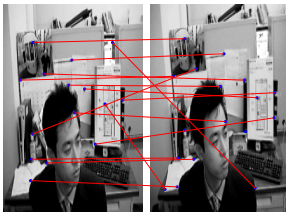


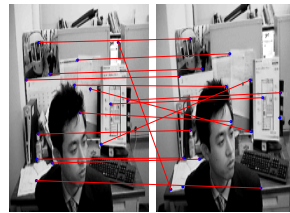
Fig. 1. Matching performance

5.2 Image Correspondences

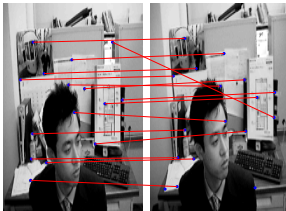
To visualize the matching for real world images we test the alternative methods on frames of video¹. We use the Harris detector to extract corner points from the first and 30th frames. We use the sum of polar sines presented in [5] to measure the similarity of every 3-tuple within the corner points and thus establish a weighted 3-uniform hypergraph for each image. Figure 2 illustrates the matching performances for alternative methods. The matching results for the two comparison methods are visualized in Figures 2(a) and 2(b), where 11 correct correspondences and 4 incorrect ones are



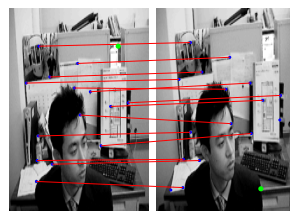
(a) PHM.



(b) TPI.



(c) DPH-DCA.



(d) DPH-DCA with two prior rejections.

Fig. 2. Image correspondences

¹ <http://www.suri.it.okayama-u.ac.jp/e-program-separate.html>

obtained by using the tensor power iteration, and 12 correct correspondences and 3 incorrect ones by the probabilistic hypergraph matching. For DCA without prior rejections (visualized in Figure 2(c)), we obtain 14 correct correspondences and 1 incorrect ones. Figure 2(d) visualizes the matching result by rejecting two outliers (green marked). It is clear that the false matching is eliminated by incorporating the proper prior rejections.

6 Conclusion and Future Work

We have presented a novel approach to high order structural matching. We have transformed the matching problem to that of extracting the dominant cluster from the direct product hypergraph for two feature sets with high order relationships. Prior knowledge about outliers can be easily involved in our framework by initializing the matchings associated with the outliers by a zero weight. Experiments have shown that our method outperforms the state-of-the-art methods.

References

1. Agarwal, S., Lim, J., Zelnik-Manor, L., Perona, P., Kriegman, D., Belongie, S.: Beyond pairwise clustering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2005)
2. Baum, L.E., Eagon, J.A.: An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society* 73, 360–363 (1967)
3. Duchenne, O., Bach, F.R., Kweon, I.S., Ponce, J.: A tensor-based algorithm for high-order graph matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2009)
4. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: Proceedings of IEEE International Conference on Computer Vision (2005)
5. Lerman, G., Whitehouse, J.T.: On d -dimensional d -semimetrics and simplex-type inequalities for high-dimensional sine functions. *Journal of Approximation Theory* 156(1), 52–81 (2009)
6. Liu, H., Latecki, L.J., Yan, S.: Robust clustering as ensembles of affinity relations. In: Proceedings of Advances in Neural Information Processing Systems (2010)
7. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(1), 167–172 (2007)
8. Rota-Bulo, S., Pelillo, M.: A game-theoretic approach to hypergraph clustering. In: Proceedings of Advances in Neural Information Processing Systems (2009)
9. Vishwanathan, S.V.N., Borgwardt, K.M., Kondor, I.R., Schraudolph, N.N.: Graph kernels. *Journal of Machine Learning Research* 11, 1201–1242 (2010)
10. Zass, R., Shashua, A.: Probabilistic graph and hypergraph matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 234–278 (2008)