A Game Theoretic Approach to Learning Shape Categories and Contextual Similarities

Aykut Erdem and Andrea Torsello

Dipartimento di Informatica
Universitá "Ca' Foscari" di Venezia
{erdem,torsello}@dsi.unive.it

Abstract. The search of a model for representing and evaluating the similarities between shapes in a perceptually coherent way is still an open issue. One reason for this is that our perception of similarities is strongly influenced by the underlying category structure. In this paper we aim at jointly learning the categories from examples and the similarity measures related to them. There is a chicken and egg dilemma here: class knowledge is required to determine perceived similarities, while the similarities are needed to extract class knowledge in an unsupervised way. The problem is addressed through a game theoretic approach which allows us to compute 2D shape categories based on a skeletal representation. The approach provides us with both the cluster information needed to extract the categories, and the relevance information needed to compute the category model and, thus, the similarities. Experiments on a database of 1000 shapes showed that the approach outperform other clustering approaches that do not make use of the underlying contextual information and provides similarities comparable with a state-of-the-art label-propagation approach which, however, cannot extract categories.

1 Introduction

The unsupervised learning of shape categories is a central problem in computer vision with significant theoretical and practical impacts. There are two interrelated aspects to the problem: The first is the discovery of the shape categories present, and this can be effectively addressed as a problem of clustering shapes, while the second is the generalization of the class properties, i.e., the ability to assign each newly encountered shape to one of the extracted classes, or to recognize it as an outlier. Fundamental to both tasks is the problem of determining how similar two shapes really are.

These issues have been extensively studied with geometric characterizations of shape using both simple descriptors such as landmark points on the boundary [4], or more complex ones such as curve descriptors [8]. Shape-classes can then be located by vectorizing the shape-attributes and applying standard central clustering techniques to the shape-vectors, while the problem of determining the membership to a class can be solved by performing principal components analysis. An alternative to the use of a single vectorial representation of the shape's geometry is to use a structural abstraction where the object is divided

into atomic components whose arrangement is then represented using a relational graph [7, 20]. Typically in this context, the similarity between two shapes is a measure of how well the primitives forming the shapes and/or their spatial organizations agree, and the assessment of whether a shape belongs to a particular class is performed by comparing in isolation the shape to one or more prototypes and by applying the nearest neighbor rule, while categories can be extracted using pairwise clustering [15].

One problem with these approaches is that they all assume the existence of a single *universal* measure of similarity between shapes, often requiring metric properties as well, while psychological experiments suggest that the human perception of similarity is not only non-metric [5], but also strongly dependent on the surrounding context [17, 11]. In particular, the observed variation within a shape-class is fundamental for determining the perception of the similarities of the shapes belonging to that class. Recently, this issue has also been surfaced from a computational point of view [19, 9].

In this paper, we propose a game theoretic approach to compute shape categories in an unsupervised way. There is a chicken and egg problem here: Class knowledge is required to determine perceived similarities, while the similarities are needed to extract class knowledge. We solve this problem using a EM-like approach where we iteratively estimate the class memberships and maximize for the parameters of our category representation. The expectation of class membership is obtained by adopting a game theoretic clustering framework presented in [16]. Then the similarities are computed as the edit-distance of a skeletal representation presented in [3] using the newly estimated cost coefficients. Central to the approach is the ability of the clustering framework to provide both the cluster information needed to extract the categories, and the relevance information (or the degree of membership) needed to compute the category model, and thus the similarities, in a robust way. Interestingly, the contextual similarity defined in [3] is not symmetric, making the ability of the game-theoretic approach to deal with asymmetric affinities particularly attractive.

2 Disconnected Skeletons and Category Influenced Matching

Skeletons are one of the most common representation scheme for generic shape recognition [20, 14], as they capture part structure and provide insensitivity to articulations and occlusions. However, in practice, two visually very similar shapes might have structurally different skeletons, hence this instability issue should be resolved either in extracting the skeleton or in the matching process. In this regard, *disconnected skeletons* [2, 1] provide an alternative solution as the method aims at obtaining a coarse yet a very stable skeleton representation from scratch.

Disconnected skeletons are defined in terms of a special distance surface (Aslan surface), the level curves of which are increasingly smoothed versions of the initial shape boundary, and which has a single extremum point that captures the center of a blob-like representation of the shape (Fig. 1(a)). Each branch extracted from this surface is classified as either *positive* or *negative*, identifying

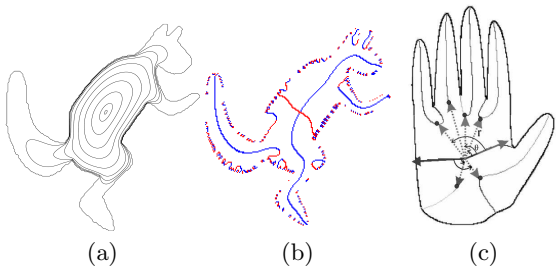

Fig. 1: Disconnected skeletons. (a) Level curves of Aslan surface (b) Positive and negative skeleton branches, respective drawn in blue and red (before pruning) (c) Spatial organization of skeleton branches (taken from [3]).

whether it originates from a positive curvature maxima (a protrusion) or a negative curvature minima (an indentation). Among the extracted branches, at least two positive and two negative branches reach the shape center, and these are called *major branches* since they represent the most prominent visual features of the shape. All the other branches terminate at some *disconnection point*s where a positive branch and a negative branch collide (Fig. 1(b)). It has been shown that these points are very stable under visual transformations such as articulation and deformation of parts. The skeletal attributes used to represent each skeleton branch are simply its type, the location of its disconnection points (r, θ), and its length l measured in the formed coordinate frame.

In [3], disconnected skeletons are represented as *rooted attributed depth-1 trees* and tree-edit distance is used to match these structures. Moreover, Baseski and coworkers [3] used the category of one of the shapes to be matched to determine the edit-costs. The cost functions are computed on the basis of category specific statistics about the skeletal attributes that are stored in an auxiliary tree union structure. In this version, the cost function for the label `change` operation is defined in terms of a generic cost function. The idea resembles Mahalanobis distance in that when the distance within the observed range of skeletal attributes, but rapidly increases outside of that region.

3 Grouping Game

In [16], a novel framework for grouping and clustering was presented which was derived from a game-theoretic formalization of the competition between the hypotheses of group membership. The basic idea is as follows: Let the hypotheses that each element belongs to a group compete with one-another, each obtaining support from compatible elements and competitive pressure from all the others. Competition will reduce the population assuming hypotheses that do not receive strong support from the rest, while it will allow populations assuming hypotheses with strong support to thrive. Eventually, all inconsistent hypotheses will be driven to extinction, while all the surviving hypotheses will reach an equilibrium with all receiving the same average support. Clustering was thus formalized as a repeated non-cooperative game where competition for class membership selects elements belonging to a coherent cluster.

Specifically, let $O = \{1, \cdots, n\}$ be the set of available elements, for each pair of strategies $i, j \in O$, a_{ij} represents the payoff of an individual playing strategy i against an opponent playing strategy j. A *mixed strategy* is a probability distribution $\mathbf{x} = (x_1, \ldots, x_n)^T$ over the available strategies O.

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n \ : \ x_i \geq 0 \text{ for all } i \in O, \ \mathbf{1}^T \mathbf{x} = 1 \right\} ,$$

where $\mathbf{1} = (1, \ldots, 1)^T$, while the *support* of a mixed strategy $\mathbf{x} \in \Delta$, denoted by $\sigma(\mathbf{x})$, is defined as the set of elements chosen with non-zero probability: $\sigma(\mathbf{x}) = \{i \in O \mid x_i > 0\}$.

The expected payoff received by a player choosing element i when playing against a player adopting a mixed strategy $\mathbf{x}$ is $(A\mathbf{x})_i = \sum_j a_{ij} x_j$, hence the expected payoff received by adopting the mixed strategy $\mathbf{y}$ against $\mathbf{x}$ is $\mathbf{y}^T A\mathbf{x}$.

The *best replies* against mixed strategy $\mathbf{x}$ is the set of mixed strategies

$$\beta(\mathbf{x}) = \{\mathbf{y} \in \Delta \mid \mathbf{y}^T A\mathbf{x} = \max_{\mathbf{z}} (\mathbf{z}^T A\mathbf{x})\} .$$

A strategy $\mathbf{x}$ is said to be a *Nash equilibrium* if it is the best reply to itself, i.e.,

$$\forall \mathbf{y} \in \Delta \quad \mathbf{x}^T A\mathbf{x} \geq \mathbf{y}^T A\mathbf{x} . \tag{1}$$

Within this formalization, Nash equilibria abstracts the main characteristics of a group: internal homogeneity, that is, a high mutual support of all elements within the group, and external dishomogeneity, or low support from elements of the group to elements that do not belong to the group. Equilibria, and thus groups, are found using the replicator dynamics [18], a well-known formalization of a natural selection process.

The main characteristics of the framework are that it is generic, as it can deal with asymmetric as well as negative affinities; it does not require *a priori* knowledge of the number of clusters as it is inherently a multi-figure/ground discrimination process; and it provides immediate measures of both the cohesiveness of the cluster in the form of its average payoff $\mathbf{x}^T A\mathbf{x}$, and of the participation of an element to the cluster. In fact the value x_i can be interpreted as a degree of participation of element i to the cluster defined by the stable point $\mathbf{x}$.

4 The Proposed Method

In this study, we attempt to solve the interrelated problems of discovering shape categories and computing the corresponding contextual similarities using a EM-like approach where we iteratively estimate the class memberships and maximize for the parameters of our category representation. The expectation of class membership is obtained by adopting the game theoretic clustering framework summarized in Section 3. Then the similarities are computed as the edit-distance of a skeletal representation presented in [3] using the newly estimated cost coefficients. The details of these steps are as follows.

4.1 Discovering Shape Categories

We define the shape category in terms of a clustering game where shapes present in the training set compete for category membership. The outcome of the competition is determined by the payoff or utility matrix $A = (a_{ij})$ which represents the similarity of shape i with shape j. Initially, these payoffs simply correspond to the similarities among the given set of shapes obtained with $a_{ij} = \exp\left(-\frac{(\mathrm{dist}(i,j))^2}{\sigma^2}\right)$ where σ is a scaling factor, and $\mathrm{dist}(i,j)$ is the tree-edit distance between the disconnected skeletons of the shapes i and j.

Since no category information is available in the beginning, the initial similarities were computed in isolation without any context, thus A is a *symmetric* matrix. However, in the subsequent iterations, the category structure discovered in the previous step influences the similarity computations by differentiating the roles of the shapes in comparison. Now, each row index corresponds to a query shape whereas each column index is a shape which has a category label assigned by the previous grouping game (if it is not found to be an outlier), and the cost functions are determined by the context about the category of the second shape. Thus, this results in an *asymmetric* similarity matrix.

Given the payoff matrix A, we extract shape categories by applying a *peal-off* strategy. At first, we start with a grouping game that considers all the shapes and we extract a cluster by running the replicator dynamics. Following that, we define a new game on the set of remaining (unlabeled) shapes and reiterate the procedure until all groups are extracted. The game theoretic framework also provides us a direct way to evaluate the coherency of extracted clusters. Let $S \in \mathcal{S}$ be an extracted group, the coherency of S can be computed as its average payoff $\mathbf{x}_S^{\mathrm{T}} A \mathbf{x}_S \in [0,1]$. By inspecting these values, we obtain an initial set $\mathcal{C}$ ($\subseteq \mathcal{S}$) of coherent shape categories which is formed by the clusters $S \in \mathcal{S}$ with $\mathbf{x}_S^{\mathrm{T}} A \mathbf{x}_S > \zeta_1$. This allows us to discard incoherent classes hence enforcing robustness in the extraction process.

The payoff information can also be used to assign additional members to the clusters in C. To compute the similarity between a shape i to a cluster S, we use the weighted similarity function $\gamma_S(i) = \frac{(A\mathbf{x}_S)_i}{\mathbf{x}_S^{\mathrm{T}} A \mathbf{x}_S}$. We evaluate this similarity measure for every unlabeled shape i and assign it to the most similar cluster if $\gamma_S(i) \geq \zeta_2$. Otherwise, it is considered as an outlier shape which does not belong to any of the extracted categories. The ability of assigning an unclustered object either to a category or to the *outlier* class is instrumental to the generalization capabilities. Note that the outlier class should be interpreted as a "don't know" label where the approach cannot say anything about the shape rather than recognizing the shape as a new class not seen in the other examples.

After reassigning the leftover elements, we re-examine the groups that were rejected by the first thresholding step and check whether they became more coherent with the removal of the reassigned elements. To evaluate their coherency we use an hysteresis strategy: we accept the groups with $|S| > 3$ whose average payoffs $\mathbf{x}_S^{\mathrm{T}} A \mathbf{x}_S > \zeta_3$, with $\zeta_3 < \zeta_1$. The purpose of this hysteresis is to reduce the effect the implicit change in scale induced by the peel-off strategy and to increase robustness with respect to the scaling factor σ.

4.2 Computing Contextual Similarities

To model the influence of the discovered category structure on the computation of shape similarities, we adopt the tree-edit distance based shape matching method proposed by Baseski *et al.* [3]. Here, however, we form the union in an unsupervised way, based on the clusters obtained with the game-theoretic approach. Further, in the computation of the edit-cost, we substitute the minimum and maximum values of the skeletal attributes in the category with soft bounds that make use of the membership information supplied by the clustering framework. In particular, we use the weighted mean $\mu_{\mathbf{x}}$ and weighted standard deviation $\sigma_{\mathbf{x}}$ (Eqn. 2) to determine the range $\mu_{\mathbf{x}} \pm 3\sigma_{\mathbf{x}}$ which has experimentally shown to account for the shape variability and provide a robust inference process.

$$\mu_{\mathbf{x}} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i}, \quad \sigma_{\mathbf{x}} = \sqrt{\frac{1}{1-\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} x_i (y_i - \mu_{\mathbf{x}})^2} \tag{2}$$

In obtaining the affinity matrix $A = (a_{ij})$ at time step $t > 0$, we introduce a soft indexing scheme where we propagate the information about the similarities to the extracted classes: When computing the similarity between the query shape i and the shape j, if j belongs to a cluster S extracted in the previous step, we multiply the similarity influenced by the new category information, with the similarity of the shape i to the cluster S normalized with respect to the most similar category. This allows us to bias the similarities towards the previously extracted clusters, thus propagating the membership throughout the iterations. Clearly, if j is an outlier shape, *i.e.* no category information is available about it, we keep the original distance which does not utilize any context information. Moreover, the corresponding multiplier b_{ij} is taken as 1.

$$a_{ij} = b_{ij} \times exp\left(-(dist(i,j))^2)/\sigma^2\right) \tag{3}$$

$$\text{where } b_{ij} = \begin{cases} 1 & \text{if } j \text{ is an outlier} \\ \dfrac{\gamma_S(i)}{\max\limits_{T \in \mathcal{C}} \gamma_T(i)} & \text{if } j \in S \end{cases}$$

Category discovery and similarity computation are iterated until the change in the ratio of unlabeled (outlier) shapes to the total number of shapes is smaller than a threshold ζ_4. Experimentally it was observed that the resulting group and distance information, as well as the query performance, are relatively stable after meeting this condition.

5 Experimental Results

In order to evaluate the performance of the proposed approach, we used the shape database provided in [3] which contains a total of 1000 shapes from 50 shape categories, each having 20 examples. We start by extracting the disconnected skeletons. After the descriptions are formed, we iteratively run the proposed

The Method	Rand Index	Corrected Rand Index	NMI
Our method at t=0	0.9818	0.9929	0.8517
Our method (asymmetric case)	0.9854	0.9933	0.8722
Normalized Cut [13] (with # of classes=51)	0.9832	0.9833	0.8381
Normalized Cut [13] (with # of classes=61)	0.9848	0.9854	0.8380
Foreground Focus [9] (with # of classes=50)	0.9748		0.7329

Table 1: The quantitative evaluation of the clustering results.

method with the empirically set parameters $\sigma^2 = 24$, $\zeta_1 = 0.85$, $\zeta_2 = 0.95$ and $\zeta_3 = 0.75$, and stop when $\zeta_4 \leq 0.005$. In this setting, the algorithm converges at the 2nd iteration. The shape categories extracted are given in Table 2, where for each class we show the shape with the highest membership score.

Table 1 shows some cluster validity measures [6] on the classes extracted with our approach. The first measure is the standard Rand index, *i.e.* the ratio of agreements over all possible pairs. The second measure is a corrected version of the Rand index where the disagreements in the outlier class are not penalized, as this class is not supposed to form a coherent group. Note that the latter form of the Rand index favors more conservative approaches, where we prefer the approach not to label a shape when in doubt, while the former version favors bolder assignments where we prefer to make a few mistakes rather than not assign a shape to a class. Which version is to be preferred is clearly dependent on the application. The last measure is the normalized mutual information (NMI) which measures the closeness between the class distributions and the ground truth.

In an attempt to assess information content in the asymmetry of the similarity matrix, we also perform the same experiment using the same parameters but rendering the affinities symmetric before applying pairwise clustering. In this case, the approach requires 3 iterations to converge. When the number of outlier shapes and the average precision recall values (Fig. 2) are considered alone, the symmetric case seems to work better than the asymmetric case. However, the difference between the plain and corrected Rand index show that the asymmetric approach is more conservative, i.e. it has a higher tendency to label shapes as belonging to an unknown class, but makes fewer misclassifications when it does assign shapes to a class, on the other hand the symmetric approach is more likely to assign shapes to a class, even when this results in more misclassifications.

We compared the results with several alternatives. The first, which should be seen as a baseline comparison, is performed by applying a pairwise clustering approach in order to extract the class structure, while assuming global, non-contextual similarities. Here we used Normalized Cut [13] as a baseline pairwise clustering approach. Note that the normalized cut approach requires the number of classes to be known *ab initio*. Here we choose two different values: 51 (the existing 50 semantic categories plus 1 for the outliers) and 61 (a number closer to the number of categories extracted with our approach). The additional number

of classes is due to the fact that there can be a substantial semantic gap between appearance and categories, and allowing more freedom can result in better overall categorization. Indeed, as it can be seen in Table 1, normalized cut performs better with more degrees of freedom, but still performs significantly worse than the proposed approach.

The second approach we are comparing against is *Foreground Focus* [9]. This is an unsupervised algorithm proposed to learn categories from sets of partially matching image features. Just like our approach, it utilizes an EM-like algorithm to infer the categories. However, its goal is to learn relevant features rather than the actual contextual similarities. In order to compare with this method, we first form *Inner-Distance Shape Context* [10] descriptions of each shape by uniformly sampling 100 landmark points across the shape boundary and using a total of 5 inner-distance bins and 12 inner-angle bins. Earth Mover's Distance (EMD) algorithm [12] is then used to compute the matchings of shape features and similarities, and Normalized Cut [13] is used to determine the clusters. Here, the total number of extracted clusters is kept fixed at 50 (the actual number of shape categories exist in the database). Table 1 and Fig. 2 show that the performance of this approach is even significantly lower than the baseline normalized cuts over the skeletal distance. The huge difference can probably be explained by the lower descriptive power of the Inner-Distance Shape Context features with respect to disconnected skeletons, or bad performance of EMD matching algorithm.

The last comparison is with the label propagation method [19] and is limited to the retrieval performance of the contextual similarities. This method has three parameters which are used to construct the affinity matrix, the neighborhood size and the window size that are respectively set as $C = 0.275$, neighborhood size $K = 10$, window size $W = 250 \times 250$. When applied to the initial (symmetric) similarities, the approach offers a slightly better precision/recall (Fig. 2). However, note that the approach solves a slightly different problem; it concentrates only on improving retrieval rate and does not provide any category structure or an estimation of perceptually relevant similarities.

6 Summary and Conclusion

In this paper, we presented an approach for the simultaneous discovering of 2D shape categories and the corresponding contextual similarities. This was achieved by adopting the game theoretic clustering approach introduced in [16] and by modifying the shape retrieval system presented in [3] in order to account for the uncertainty in the category information. The game theoretic framework naturally provides us the membership information about the extracted categories which quantifies this uncertainty, and is capable of dealing with the asymmetric similarities obtained using the contextual information. We have demonstrated the potential of the proposed framework on a large shape database composed of highly varying 1000 shapes from 50 categories.

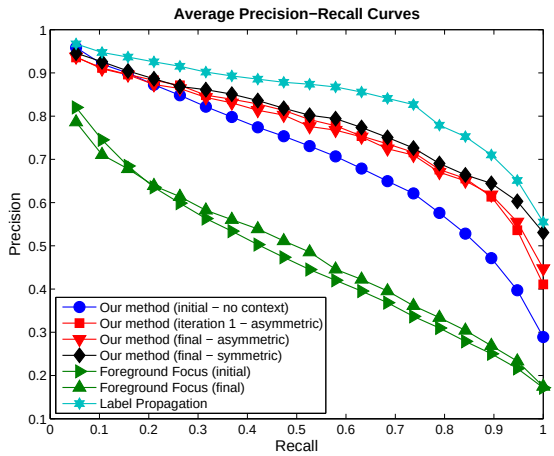

Fig. 2: Average precision-recall curves.

Acknowledgment

We acknowledge the financial support of the Future and Emerging Technology (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open project SIMBAD grant no. 213250.

References

1. C. Aslan, A. Erdem, E. Erdem, and S. Tari. Disconnected skeleton: Shape at its absolute scale. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(12):2188–2203, 2008.
2. C. Aslan and S. Tari. An axis-based representation for recognition. In *ICCV*, volume 2, pages 1339– 1346, 2005.
3. E. Baseski, A. Erdem, and S. Tari. Dissimilarity between two skeletal trees in a context. *Pattern Recognition*, 42(3):370–385, 2009.
4. T. F. Cootes, C. J. Taylor, and D. H. Cooper. Active shape models - their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995.
5. D. W. Jacobs, D. Weinshall, and Y. Gdalyahu. Classification with nonmetric distances: Image retrieval and class representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(6):583–600, 2000.
6. Anil K. Jain and Richard C. Dubes. *Algorithms for clustering data.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
7. B. B. Kimia, A. R. Tannenbaum, and S. W. Zucker. Shapes, shocks and deformations i: The components of two-dimensional shape and the reaction-diffusion space. *Int. J. Comput. Vision*, 15(3):189–224, 1995.
8. E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(3):372–383, 2004.
9. Y. J. Lee and K. Grauman. Foreground focus: Unsupervised learning from partially matching images. *International Journal of Computer Vision*, 2009.
10. H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):286–299, 2007.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Precision	1.00	0.67	1.00	0.95	1.00	0.95	1.00	1.00	0.87	0.70	0.67	1.00	1.00	0.91	0.95
Recall	0.90	0.70	0.85	0.90	0.90	0.95	0.90	0.90	0.65	0.35	0.60	0.95	0.80	1.00	1.00
Payoff	0.96	0.95	0.92	0.94	0.95	0.94	0.96	0.97	0.91	0.88	0.94	0.94	0.94	0.94	0.92
Entropy	2.75	2.77	2.69	2.72	2.81	2.77	2.77	2.89	2.47	1.60	2.75	2.63	2.39	2.94	2.87

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Precision	1.00	0.50	0.75	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Recall	0.65	0.35	0.15	1.00	0.75	0.70	0.45	0.20	0.90	0.60	0.40	0.45	0.25	0.25	0.80
Payoff	0.94	0.94	0.77	0.96	0.93	0.93	0.92	0.84	0.94	0.93	0.93	0.87	0.87	0.80	0.93
Entropy	2.39	2.55	1.34	2.92	2.59	2.46	2.03	1.38	2.78	2.44	2.03	2.13	1.61	1.57	2.71

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Precision	1.00	0.62	1.00	0.83	1.00	0.37	0.86	0.92	1.00	1.00	0.91	0.95	1.00	1.00	1.00
Recall	0.50	0.50	0.55	0.25	0.80	0.35	0.60	0.60	0.95	0.50	0.50	0.95	0.90	0.25	1.00
Payoff	0.91	0.87	0.92	0.89	0.94	0.94	0.92	0.93	0.95	0.94	0.92	0.91	0.93	0.80	0.95
Entropy	2.10	2.19	2.28	1.61	2.68	2.75	1.94	2.43	2.79	2.30	2.38	2.72	2.78	1.58	2.77

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Precision	1.00	0.75	1.00	1.00	1.00	1.00	0.74	1.00	1.00	1.00	0.95	0.94	1.00	1.00	0.55
Recall	0.80	0.15	0.95	0.90	0.35	1.00	0.70	0.75	0.90	0.60	0.95	0.75	0.20	0.85	0.55
Payoff	0.90	0.83	0.96	0.95	0.84	0.97	0.95	0.95	0.95	0.91	0.93	0.93	0.86	0.96	0.94
Entropy	2.56	1.38	2.92	2.69	1.85	2.93	2.63	2.70	2.81	2.33	2.66	2.55	1.39	2.83	2.85

	61	62	63	64
Precision	0.78	1.00	1.00	0.42
Recall	0.35	0.55	0.30	0.50
Payoff	0.90	0.92	0.85	0.91
Entropy	2.01	2.27	1.61	2.80

Table 2: The final shape categories extracted from asymmetric affinities. The number of outlier shapes is 80.

11. D. Mumford. Mathematical theories of shape: Do they model perception? In B. C. Vemuri, editor, *roc. SPIE Vol. 1570, p. 2-10, Geometric Methods in Computer Vision*, pages 2–10, 1991.
12. Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
13. Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.
14. K. Siddiqi and B.B. Kimia. A shock grammar for recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1996.
15. A. Torsello, A. Robles-Kelly, and E. R. Hancock. Discovering shape classes using tree edit-distance and pairwise clustering. *International Journal of Computer Vision*, 72(3):259–285, 2007.
16. Andrea Torsello, Samuel Rota Bulo, and Marcello Pelillo. Grouping with asymmetric affinities: A game-theoretic perspective. In *CVPR*, pages 292–299, 2006.
17. A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.
18. J. W. Weibull. *Evolutionary Game Theory.* MIT Press, Cambridge, MA., 1995.
19. X. Yang, X. Bai, L. J. Latecki, and Z. Tu. Improving shape retrieval by learning graph transduction. In *ECCV*, pages 788–801, 2008.
20. S. C. Zhu and A. L. Yuille. Forms: A flexible object recognition and modeling system. *Int. J. Comput. Vision*, 20(3):187–212, 1996.