Co-Transduction for Shape Retrieval

Xiang Bai, Bo Wang, Cong Yao, Wenyu Liu, and Zhuowen Tu

***Abstract*—In this paper, we propose a new shape/object retrieval algorithm, namely, *co-transduction*. The performance of a retrieval system is critically decided by the accuracy of adopted similarity measures (distances or metrics). In shape/object retrieval, ideally, intraclass objects should have smaller distances than interclass objects. However, it is a difficult task to design an ideal metric to account for the large intraclass variation. Different types of measures may focus on different aspects of the objects: for example, measures computed based on contours and skeletons are often complementary to each other. Our goal is to develop an algorithm to fuse different similarity measures for robust shape retrieval through a semisupervised learning framework. We name our method co-transduction, which is inspired by the co-training algorithm. Given two similarity measures and a query shape, the algorithm iteratively retrieves the most similar shapes using one measure and assigns them to a pool for the other measure to do a re-ranking, and vice versa. Using co-transduction, we achieved an improved result of 97.72% (bull's-eye measure) on the MPEG-7 data set over the state-of-the-art performance. We also present an algorithm called tri-transduction to fuse multiple-input similarities, and it achieved 99.06% on the MPEG-7 data set. Our algorithm is general, and it can be directly applied on input similarity measures/metrics; it is not limited to object shape retrieval and can be applied to other tasks for ranking/retrieval.**

***Index Terms*—Graph transduction, object retrieval, shape retrieval, similarity measure.**

I. Introduction

SHAPE-BASED object retrieval is an important task in computer vision. Given a query object, the most similar objects are retrieved from a database based on a certain similarity/distance measure, whose choice largely decides the performance of a retrieval system. Therefore, it is of critical importance to have a faithful similarity measure to account for the large intraclass and instance-level variation in configuration, nonrigid transformation, and part change. Ideally, such a similarity measure should result in smaller distances between the variants of a particular object than this object to any other ones, as well as smaller distances between intraclass objects than interclass objects. However, designing such a measure for the general retrieval task is challenging. Fig. 1 gives an illustration where a horse might have a smaller distance to a dog (based on their contours) than another horse, whereas our human vision systems can still identify them correctly.

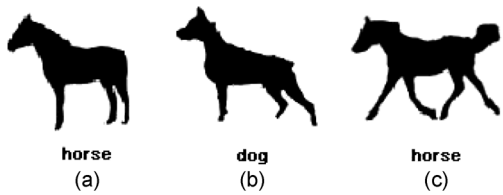


Fig. 1. A horse in (a) may look more similar to a dog in (b) than to another horse in (c).

In this paper, we refer to shape as the contour of an object silhouette. Building correspondences is often the first step in computing the shape difference, but it is challenging: Two shapes may not have direct correspondences, regardless of being represented as sparse points, closed contours, or parametric functions. For example, two shapes with the same contour but different starting points typically are considered as the same one. Therefore, measuring the similarity between two shapes often can be done in two ways: 1) by computing the direct difference in features extracted from shape contours, which are invariant to the choice of starting points and robust to a certain degree of deformation, such as moments [1] and Fourier descriptors [2]; and 2) by performing matching to find the detailed pointwise correspondences to compute the differences [3]–[8]. The latter has recently become dominant due to its ability of capturing intrinsic properties, thus leading to more accurate similarity measures.

Bai *et al.* [9] explored the group contextual information on different shapes to improve the efficiency of shape retrieval on several standard data sets [10], [11]. The basic idea was to use shapes as each other's contexts in propagation to reduce the distances between intraclass objects. The implementation was done by a graph-based transduction approach, named label propagation (LP) [12]. Later, several other graph-based transduction methods were suggested for shape retrieval [13], [14]. In addition, the method in [14] further improved the results by adding "ghost points," which were constructed based on query shape and its nearest neighbors from the database. Egozi *et al.* [15] proposed a contextual similarity function, named meta similarity, which characterizes a given object by its similarity to its k-nearest neighbor (k-NN) objects. An interesting distance learning method called contextual dissimilarity measure (CDM)

Manuscript received November 24, 2010; revised June 21, 2011; accepted September 10, 2011. Date of publication September 29, 2011; date of current version April 18, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 60903096 and Grant 60873127, by the Office of Naval Research under Grant N000140910099, and by the National Science Foundation under CAREER Award IIS-0844566. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun A. Ross.

X. Bai, C. Yao, and W. Liu are with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: xbai@hust.edu.cn; yaocong2010@gmail.com; liuwy@hust.edu.cn).

B. Wang was with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. He is now with the Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: wangbo.yunze@gmail.com).

Z. Tu is with Microsoft Research Asia, Beijing 100080, China, and also with the Laboratory of Neuro Imaging, Department of Neurology, University of California, Los Angeles, CA 90095 USA (e-mail: ztu@loni.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2170082

TABLE I
BULL'S-EYE SCORES ON THE MPEG-7 DATA SET [10] AND TARI'S DATA SET [37]. THE APPROACHES ABOVE THE SINGLE STRAIGHT LINE ARE PAIRWISE MATCHING ALGORITHMS, AND THE APPROACHES BELOW THE SINGLE STRAIGHT LINE ARE CONTEXTS-BASED ALGORITHMS

Algorithm	MPEG-7	Tari
SC [3] (DP)	86.8%	94.17%
IDSC [4]	85.4%	95.33%
DDGM [39]	80.03%	
Planar Graph Cuts [40]	85%	
Triangle Area [41]	87.3%	
Shape-tree [42]	87.7%	
ASC [17]	88.30%	95.44%
IDSC+ Perceptually Strategies [18]	88.39%	
Layered Graph [43]	88.75%	
Contour Flexibility [44]	89.31%	
AIR [8]	93.67%	
IDSC + CDM [16]	88.30%	
IDSC + LP [9]	91%	99.35%
SC + LP [9]	92.91%	97.79%
IDSC + LCDP[14]	93.32%	99.7%
SC + GM + Meta [15]	92.51%	
IDSC + Mutual Graph [13]	93.40%	
IDSC+ PS + LDCP [18]	95.60%	
ASC + LDCP [17]	95.96%	99.79%
SC + IDSC + Co-Transduction	**97.72%**	**99.995%**
IDSC + DDGM + Co-Transduction	**97.31%**	
SC + DDGM + Co-Transduction	**97.45%**	

[16] is motivated by an observation that a good ranking is usually not symmetrical in image search, which is mainly designed for the image search problem. CDM significantly improves the distance measure using bag-of-features; however, its improvement on shape retrieval is not so obvious as the shape distance measures have different properties than bag-of-features (we will show the result of CDM on a shape data set in Table I). Recently, [17] and [18] have achieved the state-of-the-art performance by using the similarity learning method [14] based on new shape similarities.

In this paper, we look at the shape retrieval problem from the "fusion" [19], [20] perspective, which is crucial for making quick and accurate decisions. Different similarity measures have different emphasis: For example, similarities computed on matching the skeletons of two objects may be robust against nonrigid transformation, but they are hard to capture the rich variability in part change; similarities computed on matching the contour parts can capture subtle change, but they may not be robust against articulation. It would be natural to fuse/combine different complementary metrics together to achieve better performance. For high-dimensional data, a direct approach of distance metric learning [21]–[23] is often used in the context of supervised learning. In another spirit, co-training style algorithms allow classifiers trained on different views [24], [25] or different subsets of the training data [26] or by different learning algorithms or parameter settings [27], [28] to pull out more samples from unlabeled data to help each other. Theoretical grounds have been studied in [24] and [29]–[31], and recently, a sufficient and necessary condition has been given, and the connection with graph-based approaches has been disclosed [32]. A straightforward way is to linearly combine a few measures together. However, this often requires a certain level of supervision or manual tuning and will not necessarily produce the best results (we will see a comparison in the experiments).

This paper provides a different way of fusing similarity/distance measures through a semisupervised learning framework, namely, *co-transduction*. The user input is a query shape, and our system returns the most similar shapes by effectively integrating two distance metrics computed by different algorithms, e.g., shape contexts (SC) [3] and inner-distance shape contexts (IDSC) [4]. Our approach is inspired by the co-training algorithm [24]. The difference, however, is that, in co-training, it requires having two conditionally independent views of the data samples. In our problem, each data only has one view, but different algorithms report measures by exploring different aspects of the data. Therefore, they may lead to different retrieval results for the same query, which can be mutual. For example, as shown in Fig. 2, the retrieval results of SC [3] in the first row and of IDSC [4] in the second row are very different as their different shape representations, although they can gain the comparable bull's-eye retrieval rate (SC: 86.8%[1]; IDSC: 85.4%) in the MPEG-7 shape data set [10].

A simple example that illustrates the motivation of the proposed method is shown in Fig. 3: In Fig. 3(1), the SC distances between query shapes A and B/C are not small due to articulation. However, in Fig. 3(2), IDSC reports a different result as it is more stable than SC for articulation changes (it uses the inner distance to replace the Euclidean distance in SC's representation). As shown in Fig. 3, the SC distance between B and C is small as they have the same pose. Although C is thicker than B, the SC distance still finds a good match between C and B. We use IDSC to retrieve B first and then put B and query A together as labeled data a new score based on the SC distance trained by A and B will give high confidence to C, as shown in Fig. 3(4). Our algorithm is inspired by co-training [24], which assumes views (sets of features) with two conditions: 1) Each view is strong enough to describe the data (a good classifier can be learned based on enough training samples); and 2) each view is conditionally independent given the labels. The pseudocode of co-training is shown in Fig. 5.

However, unlike co-training, in which two independent views (sets of features) are assumed, our algorithm deals with single-view but multiple-input similarities; we deal with the retrieval/ranking, whereas co-training is focused on the classification problem. Co-transduction is also related to [33] but with the following differences: 1) [33] tackles a regression problem; 2) k-NN was used in [33]; and 3) we focus on fusing different metrics for object retrieval. The details about the co-transduction algorithm and experiments will be given in the later sections of this paper.

II. CO-TRANSDUCTION ALGORITHM

We first briefly review the graph-based transduction algorithm (LP) [12] applied to shape retrieval [9]. Given a set of objects $X = \{x_1, \ldots, x_n\}$ and a similarity function $sim : X \times X \rightarrow R^+$ that assigns a positive similarity value to each pair of objects, assume that x_1 is a query object (e.g., a query shape) and $\{x_2, \ldots, x_n\}$ is a set of known database objects(or

[1]Here, we use dynamic programming (DP) to replace thin-plate spline as Belongie *et al.* did in [3] for the matching process and achieve 86.8% on the MEPG-7 data set. The new distance measure by DP based on the SC descriptor is used as the input for our retrieval framework.

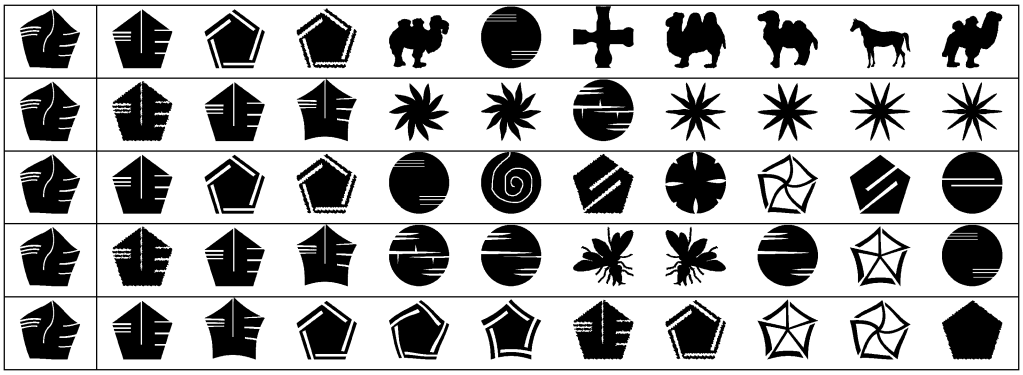


Fig. 2. The images in the first column are the query shapes. The remaining ten columns show the most similar shapes retrieved from the MPEG-7 data set. The first to fourth rows are the retrieval results of SC [3], IDSC [4], SC + LP [9], and IDSC + LP [9], respectively. The fifth row is the result of the proposed method by integrating two distance metrics computed by SC and IDSC. Note that some retrieval results by the proposed method, such as the eighth and tenth retrievals, were not retrieved by SC or IDSC since their rankings are too low with the original similarity measures.

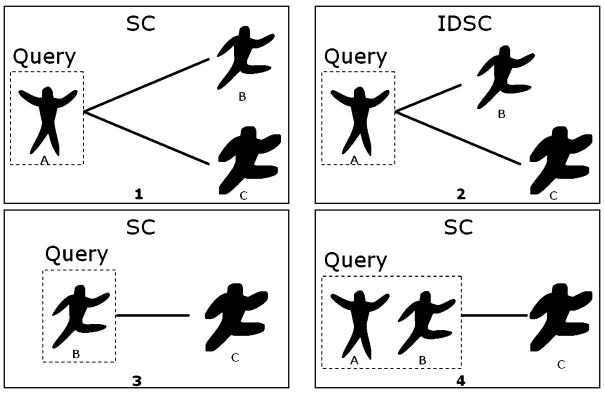


Fig. 3. Motivation of the proposed method. The shape similarities of IDSC and SC can be helpful to each other. In (1), the dissimilarities computed by SC between query shape A and database shape B/C are not very small. However, in the case of IDSC, as shown in (2), the dissimilarity between A and B is small. (3) shows that B is close to C with SC. Thus, for query A, we can use IDSC to retrieve B first and then use A and B as the query to find C by IDSC, as shown in (4).

Input: The $n \times n$ row-wise normalized similarity matrix P with the query $\{x_1 ..., x_l\}$, $f_1(x_i) = 1$ for $i = 1, ..., l$, and $f_1(x_i) = 0$ for $i = l+1, ..., n$.
while: $t < T$.
  for $i = l+1, ..., n$,
    $f_{t+1}(x_i) = \sum_{j=1}^{n} P_{ij}\ f_t(x_j)$
  end
  $f_{t+1}(x_i) = 1$ for $i = 1, ..., l$.
end
Output: The learned new similarity values to the query $\{x_1, ...x_l\}$: f_T.

Fig. 4. Pseudocode of the LP algorithm when the query includes a group of objects. T is the iteration number of LP, which is fixed in our experiments.

a training set). Then, by sorting the values $sim(x_1, x_i)$ in decreasing order for $i = 2, \ldots, n$, we can obtain a ranking for database objects according to their similarity to the query. A critical issue is then to learn a faithful sim. Bai *et al.* [9] applied LP to learn a new similarity function sim_T that drastically improves the retrieval results of sim for the given query x_1. They let $w_{i,j} = sim(x_i, x_j)$, for $i, j = 1, \ldots, n$, be a similarity matrix, then obtain an $n \times n$ probabilistic transition matrix P as a row-wise normalized matrix w, i.e.,

$$P_{ij} = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}} \tag{1}$$

where P_{ij} is the probability of transit from node i to node j.

A new similarity measure s is computed based on P. Since s is defined as the similarity of other elements to query x_1, we denote $f(x_i) = s(x_1, x_i)$ for $i = 1, \ldots, n$. A key function is f, and it satisfies

$$f(x_i) = \sum_{j=1}^{n} P_{ij} f(x_j). \tag{2}$$

Thus, the similarity of x_i to query x_1, which is expressed as $f(x_i)$, is a weighted average over all other database objects, where the weights sum to 1 and are proportional to the similarity of the other database objects to x_i. In other words, a function $f : X \rightarrow [0, 1]$ such that $f(x_i)$ is a weighted average of $f(x_j)$, where the weights are based on original similarities $w_{i,j} = sim(x_i, x_j)$.

Note that LP is not limited to only one query object, which can be also used for two or more queries as it is a classification method (see the case in Fig. 3(4), where there are two query objects A and B). Assume that $\{x_1, \ldots, x_l\}$ is a group of query objects and $\{x_{l+1}, \ldots, x_n\}$ is a set of known database objects. Then, the LP algorithm for computing the new similarity is shown in Fig. 4.

In a general situation, graph-based transduction can be viewed as performing manifold regularization [34]. $f^* = \arg\min_{f \in \mathcal{H}_K} \sum_{i=j}^{l} V(x_j, y_j, f) + \lambda_1 \|f\|^2_{\mathcal{H}_K} + \lambda_2 \mathbf{f}^T L$, which is an approximation to the continuous function space of f based on the labeled (query objects in our case) and unlabeled data (database objects). L is the Laplacian map computed from similarity measures P. $V(x_j, y_j, f)$ measures the classification error of f on the supervised data, and $\|f\|^2_{\mathcal{H}_K}$ is a regularized form of f. Now, we view LP as a tool for improving an input similarity function by taking the contextual information between objects. The key problem we want to address in this paper is how to build a robust retrieval system

Input: the labeled training set L
the unlabeled training set U
Process:
Create a pool U' of examples by choosing u examples at random from U
Loop for k iterations:
Use L to train a classifier h_1 that considers only the x_1 portion of x
Use L to train a classifier h_2 that considers only the x_2 portion of x
Allow h_1 to label p positive and n negative examples from U'
Allow h_2 to label p positive and n negative examples from U'
Add these self-labeled examples to L
Randomly choose $2p + 2n$ examples from U to replenish U'

Fig. 5. Co-training algorithm by Blum and Mitchell [24].

Input: a query object x_1 (a labeled data)
the database objects $X = \{x_2, ...x_n\}$ (unlabeled data)
Process:
Create a $n \times n$ probabilistic transition matrix P_1 based on one type of shape similarity (eg. SC)
Create a $n \times n$ probabilistic transition matrix P_2 based on another type of shape similarity (eg. IDSC)
Create two sets Y_1, Y_2 such that $Y_1 = Y_2 = \{x_1\}$
Create two sets X_1, X_2 such that $X_1 = X_2 = X$
Loop for m iterations:
Use P_1 to learn a new similarity sim_1^j by graph transduction when Y_1 is used as the query objects ($j = 1, ..., m$ is the iteration index)
Use P_2 to learn a new similarity sim_2^j by graph transduction when Y_2 is used as the query objects
Add the p nearest neighbors from X_1 to Y_1 based on the similarity sim_1^j to Y_2
Add the p nearest neighbors from X_2 to Y_2 based on the similarity sim_2^j to Y_1
$X_1 = X_1 - Y_1$
$X_2 = X_2 - Y_2$
(Then X_1, X_2 will be unlabeled data for graph transduction in the next iteration)

Fig. 6. Co-transduction algorithm.

given two (multiple) input similarity measures. A straightforward solution is to linearly combine different measures and use LP to gain further improvement. We will later show that this yields less encouraging results than the proposed algorithm, which is co-transduction.

Figs. 5 and 6 give the pseudocodes for co-training [24] and the proposed co-transduction algorithm, respectively. Same as in Bai *et al.* [9], a query object x_1 and database objects $\{x_2, \cdot, x_n\}$ are, respectively, considered as labeled and unlabeled data for graph transduction. In spirit, co-transduction is in the co-training family; unlike the original co-training algorithm, co-transduction emphasizes single-view but different metrics. It uses one metric to pull out confident data for the other metric to refine the performance. In implementation, the nearest neighbors of the query object are added to the labeled data set for graph transduction in the next iteration based on the other shape similarity. The final similarity sim_F of co-transduction is the average of all the similarities, i.e., $sim_F = (1/2m)\sum_{j=1}^{m}(sim_1^j + sim_2^j)$.

When the database of known objects is large, computing all n objects becomes impractical; in practice, we construct similarity matrix w using the first $M \ll n$ most similar objects to query x_1 according to the original similarity, which is similar to Bai *et al.* [9]. Let S denote the first M similar objects to query x_1. As different shape similarities often have different S, we use S_1 and S_2 to represent the first M similar objects to x_1 according to two kinds of shape similarity, respectively. Then, the pseudocode of an efficient version of the co-transduction algorithm is shown in Fig. 7, which is used in all our experiments. In our experiments, M is always set to 300.

A. Theoretical Justification

Next, we provide a brief theoretical discussion of our algorithm. We borrow the analysis from [31], which mostly follows the probably approximately correct (PAC) learning theory. Let H_1^0 and H_2^0 be two classifiers (the two transduction algorithms on different metrics in our case) at round 0. They are, respectively, bounded by generalization errors $a_0 < 0.5$ and

Input: a query object x_1 (a labeled data)
the database objects $X = \{x_2, ...x_n\}$ (unlabeled data)
Process:
Create a $M \times M$ probabilistic transition matrix P_1 based on one type of shape similarity with the data from S_1
Create a $M \times M$ probabilistic transition matrix P_2 based on another type of shape similarity with the data from S_2
Create two sets Y_1, Y_2 such that $Y_1 = Y_2 = \{x_1\}$
Create two sets X_1, X_2 such that $X_1 = X_2 = X$
Loop for m iterations:
Use P_1 to learn a new similarity sim_1^j by graph transduction when Y_1 is used as the query objects ($j = 1, ..., m$ is the iteration index)
Use P_2 to learn a new similarity sim_2^j by graph transduction when Y_2 is used as the query objects
Add $N_1 \bigcap S_2$ (N_1 denotes the p nearest neighbors from X_1 to Y_1 based on the similarity sim_1^j) to Y_2
Add $N_2 \bigcap S_1$ (N_2 denotes the p nearest neighbors from X_2 to Y_2 based on the similarity sim_2^j) to Y_1
$X_1 = X_1 - Y_1$
$X_2 = X_2 - Y_2$
(Then X_1, X_2 will be unlabeled data for graph transduction in the next iteration)

Fig. 7. Co-transduction algorithm for a large database.

$b_0 < 0.5$ with a high probability, i.e., $1-\delta$, in PAC. Then, H_1^0 selects u number of unlabeled data samples (database objects) and put them into data set σ_2, which contains all the examples for training H_2^1 using transduction, where H_2^1 denotes the learned classifier based on σ_2 after the first round in our case. Let l be the number of labeled data and $G = u \times a_0$. If $l \times b_0 \leq e \sqrt[G]{G!} - G$, then

$$Pr\left[d\left(H_2^1, H^*\right) \geq b_1\right] \leq \delta$$

where H^* is the ideal classifier to retrieve all the correct answers, and $d(H_2^1, H^*)$ measures the difference between learned H_2^1 and H^*. The new error is then [31]

$$b_1 = \max\left[\frac{l \times b_0 + u \times a_0 - u \times d\left(H_1^0, H_2^1\right)}{l}, 0\right].$$

Here, l is the number of labeled examples, and u is the number of unlabeled data. In order to achieve good classifiers whose generalization errors are less than b_1, sampled sequence σ_2 must be sufficient to guarantee that no classifier whose generalization error is no smaller than b_1 has a lower observed rate of disagreement with σ_2 than H^* with a probability greater than $1-\delta$. As we can see, the general guidance to achieve a small b_1 is to reduce the errors of the original learners (good input metrics) and increase the complementariness of the metrics. Our algorithm does not necessarily improve the overall performance if the input metrics are not so good at the first place and they are not so different from each other. Analogously, by minimizing the empirical risk, we can obtain the classifier that has the lowest observed rate of disagreement with sampled sequence σ_2.

From a different perspective, different measures explore different aspects about similarity; the top M most similar objects with respect to each measure are often not all correct; however, the most similar one (nearest neighbor) is likely be the case; pulling out the best match by one measure to the other helps further retrieve similar ones by the other complementary measures. This intuition explains why co-transduction works. Our work is also related to the diffusion map [35], which obtains improved similarity measures for clustering by performing Markov random walks. Our transductive learning component improves similarity measures just like the diffusion map algorithm, and the fusion of different metrics gives further improvement. By exchanging the improved similarity measures of two transductive learning algorithms, we gradually achieve a fused similarity by letting two originally different measures meet with each other, which realizes a fusion process.

III. Extend Co-Transduction to Tri-Transduction

In the co-transduction algorithm, we only provide a solution of combining two kinds of similarities. Here, we proposed an algorithm called tri-transduction, which can be used for combining three kinds of similarities. We are inspired by [36], which combines three classifiers for improving the classification accuracy. We show the details of the tri-transduction algorithm in Fig. 8. As shown in Fig. 8, the spirit of tri-transduction is very similar to co-transduction. Assume that A, B, and C are the input three kinds of similarities of tri-transduction, tri-transduction retrieves the most similar object by similarity A and add them to the pool of B to do a re-ranking. Meanwhile, the most similar objects by similarity B is added to C, and the most similar object by C is added to A. In such a way, an iterative procedure can be realized to enhance each input similarity. Same as that in co-transduction, the final similarity of tri-transduction is the average of all these similarities.

Input: a query object x_1 (a labeled data)
the database objects $X = \{x_2, ...x_n\}$ (unlabeled data)
Process:
Create a $n \times n$ probabilistic transition matrix P_1 based on the first type of shape similarity
Create a $n \times n$ probabilistic transition matrix P_2 based on the second type of shape similarity
Create a $n \times n$ probabilistic transition matrix P_3 based on the third type of shape similarity
Create two sets Y_1, Y_2, Y_3 such that $Y_1 = Y_2 = Y_3 = \{x_1\}$
Create two sets X_1, X_2, Y_3 such that $X_1 = X_2 = X_3 = X$
Loop for m iterations:
Use P_1 to learn a new similarity sim_1^j by graph transduction when Y_1 is used as the query objects ($j = 1, ..., m$ is the iteration index)
Use P_2 to learn a new similarity sim_2^j by graph transduction when Y_2 is used as the query objects
Use P_3 to learn a new similarity sim_3^j by graph transduction when Y_3 is used as the query objects
Add the p nearest neighbors from X_1 to Y_1 based on the similarity sim_1^j to Y_2
Add the p nearest neighbors from X_2 to Y_2 based on the similarity sim_2^j to Y_3
Add the p nearest neighbors from X_3 to Y_3 based on the similarity sim_3^j to Y_1
$X_1 = X_1 - Y_1$
$X_2 = X_2 - Y_2$
$X_3 = X_3 - Y_3$
(Then X_1, X_2, X_3 will be unlabeled data for graph transduction in the next iteration)

Fig. 8. Tri-transduction algorithm.

IV. Experimental Results

Here, we show results on three data sets, namely, MPEG-7 shape [10], Tari's shape [37], and Wei's trademark [38]. In addition, we show that our algorithm can potentially benefit from bag-of-word-based image search.

A. Results on Shape Data sets

The MPEG-7 shape data set consists of 1400 silhouette images grouped into 70 classes, with each class having 20 different shapes. Usually, the retrieval rate for this data set is measured by the "bull's-eye test." Every shape in the database is compared with all other shapes, and the number of shapes from the same class among the 40 most similar shapes is reported. The bull's-eye retrieval rate is the ratio of the total number of shapes from the same class to the possible number (which is 20×1400). We use the similarities computed by SC [3] and IDSC [4] as the input distance measures. The new similarity obtained by co-transduction results in 97.72% (bull's-eyes), which outperforms existing state-of-the-art algorithms; to further show our algorithm being a general method, we also use the similarity computed by Tu and Yuille [39] together with SC and IDSC as the distance measures for co-transduction and achieve scores of 97.45% and 97.31%, respectively. These improvements show the effectiveness of our algorithm. Our results and those by several state-of-the-art methods on the MPEG-7 data set are shown in Table I. We observe that co-transduction outperforms the alternatives. This demonstrates that integrating different shape similarities is an important direction for shape recognition. Notice that the state-of-the-art result without LP/diffusion on the MPEG-7 data set is achieved by articulation-invariant representation (AIR) [8], and we believe that co-transduction can achieve a higher score when using the similarity computed by AIR as the input.

In order to visualize the gain in retrieval rates by our method compared with SC or IDSC, we plot the percentage of correct results among the first k most similar shapes in Fig. 9(a). For example, we plot the percentage of the shapes from the same class among the first k-NNs for $k = 1, \ldots, 40$. Recall that each class has 20 shapes, and this is the reason for curve $k > 20$. We observe that not only does the proposed method increase the bull's-eye score but also the ranking of the shapes for all $k = 1, \ldots, 40$ gets improved. In Fig. 9(a), we also plot the curves of retrieval rates for SC/IDSC with graph transduction [9] (e.g., SC + LP and IDSC + LP).

Tari's data set [37] consists of 1000 silhouette images grouped into 50 classes, with 20 images per class. Tari's data set has more articulation changes within each class than the MPEG-7 data set, as shown in Fig. 10, and consequently, IDSC

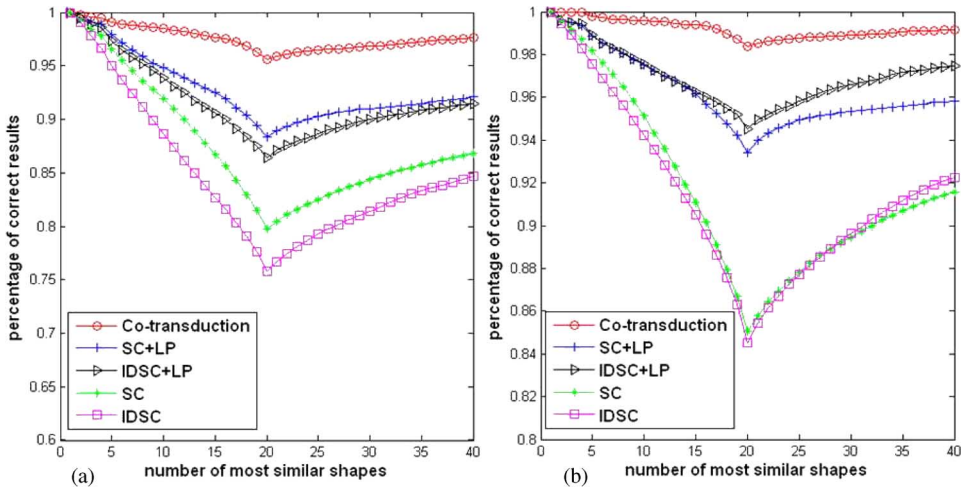


Fig. 9. Curves of retrieval rates for SC, IDSC, SC + LP, IDSC + LP, and co-transduction on the (a) MPEG-7 shape data set and (b) Tari's data set (b).

Fig. 10. Sample images in Tari's data set.

achieved better results than SC on this data set (see Table I). The retrieval performance on this data set is also measured by the "bull's-eye test." Only one error was made when retrieving all the shapes from the data set, with the nearly perfect retrieval rate of 99.995%. Table I also lists several results of Tari's data set in comparison with other approaches; we observe that the second highest result by ASC + LCDP [17] is 99.79%, with 42 errors. Same as in Fig. 9(a), the retrieval curves in Fig. 9(b) are plotted to clearly show the performance gain by the co-transduction algorithm.

In addition to shape retrieval, co-transduction can be also used to improve the performance of shape recognition. The accuracy of shape recognition is usually measure by 1-nearest neighbor classification, which means the object is simply assigned to the class of its nearest neighbor. Table II lists several recent state-of-the-art methods and their recognition rates on the MPEG-7 data set. Note that the methods [45]–[47] are tested in a different manner: Before testing, a multiclass classifier is trained based on features computed from training shapes. They adopt the leave-one-out strategy for the supervised training, which means 19 shapes from each class of the MPEG-7 data set are used for training and the remaining one is used for testing. Our algorithm significantly outperforms these methods without knowing the labels of the database objects.

In addition to the MPEG-7 and Tari's data sets, several other data sets, such as Kimia's [11] and Rutgers tool [6], are also very popular for shape retrieval. Since the sizes of these data sets are relatively small and graph transduction [9] has achieved nearly perfect results on these data sets, we do not use them here.

B. Results on Trademark Images

We also tested our method on a trademark data set [38] consisting of 14 different classes with 1003 trademark images.

TABLE II
SHAPE RECOGNITION RATES ON MPEG-7 DATA SETS

Method	Classification Accuracy
Chance Probabilities [48]	97.4%
Class Segment Sets [45]	97.93%
Polygonal Multi-resolution [49]	97.57%
Robust Symbolic [50]	98.57%
Integrating Contour and Skeleton [46]	98.93%
Kernel Edit Distance [47]	98.93%
IDSC [4]	98.00%
SC [3](DP)	98.64%
IDSC + LP [9]	98.79%
SC + LP [9]	99.00%
SC+IDSC+Co-Transduction	**99.50%**

Fig. 11 shows some typical examples from this trademark data set. We use the precision–recall curve for the evaluation. The x- and y-axes represent recall and precision rates, respectively. Precision is the ratio of the number of relevant images retrieved to the total number of images retrieved, whereas recall is ratio of the number of relevant images to the total number of relevant images stored in the database. For each query image input to the system, the system returns 11 pages of hits with descending similarity rankings, with each page containing nine trademark images. This allows the performance of our system to be evaluated on a pagewise manner. Since there are only five classes containing more than 99 images, we only report the precision–recall graph on these five classes. Each curve consists of 11 data points, with the ith point from the left corresponding to the performance when the first i pages of hits are taken into consideration. A precision–recall line stretching longer horizontally and staying high in the graph indicates good performance. Here, we use two distance measures, namely, moment invariants [51] and Zernike [52]. In Fig. 12, the data points shown on the curve for co-transduction are the average

Fig. 11. Sample images in Wei's trademark data set.

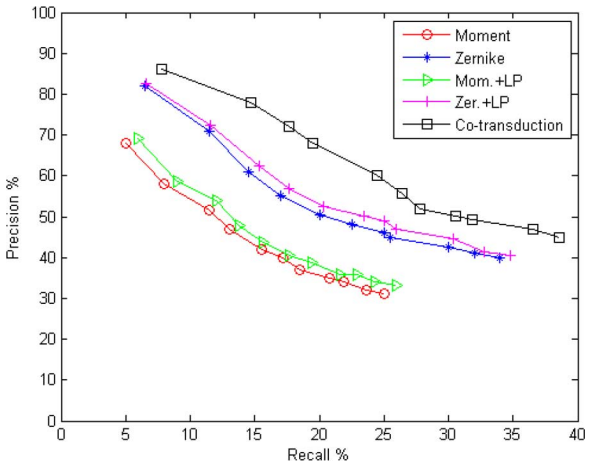


Fig. 12. Precision–recall curves for trademark images.

Fig. 13. Sample images of the N–S data set [55].

precision and recall rates over the five classes. The curves show that our method can improve the performance of trademark retrieval significantly.

C. Improving Bag-of-Features Image Search With Co-Transduction

Here, we show that co-transduction can be used to improve the accuracy of image search. Bag-of-features image representation [53], [54] is widely used in image search. Recently, Jegou *et al.* [16] have proposed a distance learning method called CDM. We compare our method with CDM on the Nistér and Stewénius (N–S) data set [55]. The N–S data set consists of 2550 objects or scenes, each of which takes four different viewpoints. Hence, the data set has 10 200 images in total. A few example images from the N–S data set are shown in Fig. 13.

We adopt the method in [16] to compute the similarity for image search. The image descriptor is a combination of the Hessian–Affine region detector [56] and the SIFT descriptor [57]. A visual vocabulary is obtained using the k-means algorithm on the subsampled image descriptors. As co-transduction requires two input similarity measures, we propose another similarity, named reverse similarity, based on the one by [16]. Let $w_{i,j}$ denote the similarity between objects i and j computed by [16], reverse similarity $w^r_{i,j} = 1/d^\beta$, where d is the ranking number of i when using j as a query for the data set, and β is a weight factor setting with a constant 10. Reverse similarity is motivated by the phenomenon pointed out by [16]: A good ranking is usually not symmetrical in image search, which tells us that two objects can be very likely from the same category when they both obtain a good ranking position when using each other as a query. With w and w^r, we can apply co-transduction to image search on the N–S data set, and the measure score is the average number of correct images among the first four images returned. Table III lists the results on the N–S data set. We observe that co-transduction significantly increases the score from 3.26 to 3.66, which is also better than CDM's result when the number visual vocabulary is 1 and the vocabulary size is 30 000. Our result demonstrates that co-transduction can be also applied to image search for performance enhancement.

D. Improving Shape Retrieval With Tri-Transduction

The retrieval power of co-transduction has been shown by several experiments in this paper. Here, we test the tri-transduction algorithm on the MPEG-7 shape data set with three kinds of similarities, namely, SC [3], IDSC [4], and DDGM [39]. One problem of tri-transduction is that the order of the input three similarities will impact the final performance, as the three kinds of similarities have different discriminatory powers. Thus, we have two kinds of choices for the input order: SC– IDSC–DDGM and SC–DDGM–IDSC (The other orders, in fact, are the same with these two). We list the bull's-eye scores of tri-transduction on the MPEG-7 data set with the above input orders in Table IV. We observe that tri-transduction outperforms co-transduction (see the results in Table I) and achieves a record-breaking performance.

E. Experiments on the Unbalanced Shape Data Sets

Here, we evaluate the performance of co-transduction on unbalanced shape data. Since the number of instances is the same for each class in almost all the shape benchmarks, we randomly divide the MPEG-7 shape data set into five groups, namely, G1, G2, G3, G4, and G5, and each group contains 14 classes. We remove some shapes from each group to make the data set unbalanced in the following manner: For each class from the different groups, we keep different numbers of instances. Specifically, the numbers for each class in G1, ..., G5 are 20, 16, 12, 8, and 4, respectively. We generate ten different unbalanced shape data sets with the above strategy and still adopt bull's-eye scores to evaluate the performance of shape retrieval. The results listed in Table V are the average of retrieval scores on the ten unbalanced data sets. These results demonstrate that co-transduction can still achieve a significant improvement although the shape data sets are unbalanced.

F. Parameter Setting and Time Complexity Analysis

As introduced in [9], there are two key parameters for LP, i.e., α and K. In addition to α and K, there are two additional parameters for co-transduction, i.e., iteration number m and number of nearest neighbors p. For the MPEG-7 and Tari's data sets, we use the following parameter settings: $\alpha = 0.25$, $K = 14$, (which are consistent with the settings in [9]), $m = 4$, and $p = 3$. The parameter settings of tri-transduction (the experiments in Section IV-D) are the same with co-transduction. For the trademark data set, since the input distance measures are different from the ones for MEPG-7 data set, the parameter setting

TABLE III
RESULTS ON THE N–S DATA SET

number of distinct visual vocab.	vocab. size	original N-S score	N-S score with CDM	N-S score with co-transduction
1	30000	3.26	3.57	3.66

TABLE IV
RESULTS ON THE MPEG-7 SHAPE DATA SET WITH TRI-TRANSDUCTION

Input similarities	SC, IDSC, DDGM	SC, DDGM, IDSC
Bull eyes scores	99.06%	99.04%

TABLE V
RESULTS ON UNBALANCED SHAPE DATA SETS

Method	Result
SC	87.14%
IDSC	84.63%
SC + LP	92.70%
IDSC + LP	92.19%
SC + IDSC + Co-Transduction	97.40%

TABLE VI
BULL'S-EYE SCORES ON THE MEPG-7 DATA SET WITH DIFFERENT PARAMETER SETTINGS

	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
$p = 1$	96.89%	97.05%	97.30%	97.32%	97.34%
$p = 2$	97.06%	97.24%	97.36%	97.45%	97.36%
$p = 3$	97.20%	97.54%	97.63%	97.72%	97.67%
$p = 4$	97.13%	97.30%	97.42%	97.37%	97.32%
$p = 5$	97.24%	97.55%	97.58%	97.20%	96.92%

is $\alpha = 8$, $K = 8$, $m = 2$, and $p = 2$. For the N–S data set, the parameters are $\alpha = 0.25$, $K = 10$, $m = 3$, and $p = 1$. Since [9] has introduced a supervised learning method for determining parameters α and K in detail, we no longer review it here. We only need to focus on m and p. As both m and p are integers, their values are very easy to set. Table VI shows the scores on the MPEG-7 data set when setting m and p with integers from 1 to 5. We observe that all these scores are around 97%, which demonstrates the insensitiveness of co-transduction to parameter tuning.

To further justify the effectiveness of the co-transduction method, we iteratively run LP on the MPEG-7 data set based on only one type of similarity with the same parameter setting for co-transduction (the p most similar objects will be added into the query set for the next iteration), and we get the bull's-eye scores of 92.68% and 91.79% based on SC and IDSC, respectively. Compared with LP's results in Table I, there is not so much change. Let sim'_{SC} and sim'_{IDSC} denote the similarities obtained in the above experiments. We obtain a new similarity sim'_c by linearly combining sim'_{SC} and sim'_{IDSC} as follows: $sim'_c = \lambda sim'_{\mathrm{SC}} + (1 - \lambda) sim'_{\mathrm{IDSC}}$, where λ is a weight factor. We tuned λ, and the highest score based on sim'_c is 92.0% when λ is 0.9. These scores are much lower than the ones by co-transduction, and this illustrates that a direct linear combination is not always desirable.

The time complexity for one iteration of the LP algorithm is $O(n^2)$, where n is the number of the database objects. As aforementioned, in our implementation, we use only the first $M \ll n$ most similar objects to the query object to construct the similarity matrix for LP. Thus, the complexity for each iteration of LP is $O(M^2)$. The whole complexity of LP is $O(M^2T)$, where T is the iteration number. It is easy to know that the proposed co-transduction/tri-transduction algorithm is $O(M^2Tm)$, as they adopt the LP algorithm with m iterations. This complexity is acceptable since only T is a large number. (In our experiments, $M = 300$ and $T = 5000$, which are both fixed. m is often smaller than 10.) Specifically, for the MPEG-7 data set, the average time of co-transduction for one retrieval is about 9 s on a common personal computer with a 2.53-GHz central processing unit.

V. LIMITATION OF THE PROPOSED ALGORITHM

Although our algorithm achieves encouraging results on several large benchmarks, it may fail in some cases. In particular, when the input distance measures are quite different from the truth or the variance of data from the same class is too large, our algorithm will not lead to any improvement in performance. For example, our algorithm did not improve the retrieval rates on labeled face in the wild (LFW) [58]. LFW is a very challenging data set for face recognition since it contains 13 233 face images collected from *Yahoo! News* in 2002–2003 and the faces in it show a big variety in lighting, pose, appearance, etc. We use SIFT [57] and LBP [59] as the input distance measures for co-transduction, and the result obtained by the learned distance with the proposed algorithm is almost the same with the ones by the original distances.

VI. CONCLUSION

We have proposed a shape retrieval framework, named co-transduction, which combines two different distance metrics. With the same spirit as co-transduction, tri-transduction combines three different distance metrics. The significant performance improvement on four large data sets has demonstrated the effectiveness of co-transduction/tri-transduction for shape/object retrieval. Our future work includes extending to other problems and providing deeper understanding of the approach.

ACKNOWLEDGMENT

The authors would like to thank the four anonymous reviewers for their valuable suggestions.

REFERENCES

[1] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
[2] C. Zahn and R. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 269–281, Mar. 1972.
[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[4] H. Ling and D. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007.

[5] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, no. 2/3, pp. 114–141, Mar. 2003.

[6] K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker, "Shock graph and shape matching," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 13–32, Nov. 1999.

[7] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1274–1286, Oct. 2003.

[8] R. Gopalan, P. K. Turaga, and R. Chellappa, "Articulation-invariant representation of non-planar shapes," in *Proc. ECCV*, 2010, pp. 286–299.

[9] X. Bai, X. Yang, L. Latecki, W. Liu, and Z. Tu, "Learning context sensitive shape similarity by graph transduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 861–874, May 2010.

[10] L. Latecki, R. Lakámper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proc. CVPR*, 2000, pp. 424–429.

[11] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 550–571, May 2004.

[12] X. Zhu, "Semi-supervised learning with graphs," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 2005, CMU–LTI–05–192.

[13] P. Kontschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis," in *Proc. ACCV*, 2009, pp. 655–666.

[14] X. Yang, S. Koknar-Tezel, and L. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *Proc. CVPR*, 2009, pp. 357–364.

[15] A. Egozi, Y. Keller, and H. Guterman, "Improving shape retrieval by spectral matching and meta similarity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1319–1327, May 2010.

[16] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 2–11, Jan. 2010.

[17] H. Ling, X. Yang, and L. Latecki, "Balancing deformability and discriminability for shape matching," in *Proc. ECCV*, 2010, pp. 411–424.

[18] A. Temlyakov, B. Munsell, J. Waggoner, and S. Wang, "Two perceptually motivated strategies for shape classification," in *Proc. CVPR*, 2010, pp. 2289–2296.

[19] H. B. Mitchell, *Multi-Sensor Data Fusion? An Introduction*. Berlin, Germany: Springer-Verlag, 2007.

[20] S. Das, *High-Level Data Fusion*. Norwood, MA: Artech House, 2008.

[21] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. J. Russell, "Distance metric learning with application to clustering with side-information," in *Proc. NIPS 15*, 2002, pp. 505–512.

[22] M. Schultz and T. Joachims, "Learning a distance metric from relative comparisons," in *Proc. NIPS 16*, 2004, pp. 41–48.

[23] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. ICML*, 2007, pp. 209–216.

[24] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. COLT*, 1998, pp. 92–100.

[25] V. Sindhwani, P. Niyogi, and M. Belkin, "A co-regularization approach to semi-supervised learning with multiple views," in *Proc. ICML*, 2005, pp. 824–831.

[26] Z. Zhou and M. Li, "Tri-training: Exploit unlabeled data using three classifiers," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 11, pp. 1529–1541, Nov. 2005.

[27] S. Goldman and Y. Zhou, "Enhancing supervised learning with unlabeled data," in *Proc. ICML*, 2000, pp. 327–334.

[28] Z. Zhou and M. Li, "Semi-supervised regression with co-training," in *Proc. IJCAI*, 2005, pp. 908–913.

[29] S. Dasgupta, M. Littman, and D. McAllester, "PAC generalization bounds for co-training," in *Proc. NIPS 14*, 2002, pp. 375–382.

[30] M.-F. Balcan, A. Blum, and K. Yang, "Co-training and expansion: Towards bridging theory and practice," in *Proc. NIPS 17*, 2005, pp. 89–96.

[31] W. Wang and Z.-H. Zhou, "Analyzing co-training style algorithms," in *Proc. ECML*, 2007, pp. 454–465.

[32] W. Wang and Z.-H. Zhou, "A new analysis of co-training," in *Proc. ICML*, 2010, pp. 1135–1142.

[33] Z.-H. Zhou and M. Li, "Semi-supervised regression with co-training," in *Proc. IJCAI*, 2004.

[34] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, 2006.

[35] R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, 2006.

[36] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 11, pp. 1529–1541, Nov. 2005.

[37] C. Aslan, A. Erdem, E. Erdem, and S. Tari, "Disconnected skeleton: Shape at its absolute scale," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2188–2203, Dec. 2008.

[38] C.-H. Wei, Y. Li, W. Y. Chau, and C.-T. Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure," *Pattern Recognit.*, vol. 42, no. 3, pp. 386–394, Mar. 2009.

[39] Z. Tu and A. L. Yuille, "Shape matching and recognition—Using generative models and informative features," in *Proc. ECCV*, 2004, pp. 195–209.

[40] F. R. Schmidt, E. Toeppe, and D. Cremers, "Efficient planar graph cuts with applications in computer vision," in *Proc. CVPR*, 2009, pp. 351–356.

[41] N. Alajlan, M. Kamel, and G. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1003–1013, Jun. 2008.

[42] P. F. Felzenszwalb and J. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. CVPR*, 2007, pp. 1–8.

[43] L. Lin, K. Zeng, X. Liu, and S. Zhu, "Layered graph matching by composite cluster sampling with collaborative and competitive interactions," in *Proc. CVPR*, 2009, pp. 1351–1358.

[44] C. Xu, J. Liu, and X. Tang, "2D shape matching by contour flexibility," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 180–186, Jan. 2009.

[45] K. Sun and B. Super, "Classification of contour shapes using class segment sets," in *Proc. CVPR*, 2005, pp. 727–733.

[46] X. Bai, W. Liu, and Z. Tu, "Integrating contour and skeleton for shape classification," in *Proc. IEEE Workshop NORDIA*, 2009, pp. 360–367.

[47] M. Daliri and V. Torre, "Shape recognition based on kernel edit distance," *Comput. Vis. Image Understand.*, vol. 114, no. 10, pp. 1097–1103, Oct. 2010.

[48] B. Super, "Retrieval from shape databases using chance probability functions and fixed correspondence," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 20, no. 8, pp. 1117–1137, Dec. 2006.

[49] E. Attalla and P. Siy, "Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching," *Pattern Recognit.*, vol. 38, no. 12, pp. 2229–2241, Dec. 2005.

[50] M. Daliri and V. Torre, "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognit.*, vol. 41, no. 5, pp. 1799–1815, May 2008.

[51] R. Gonzalez, R. Woods, and S. Eddins, *Digital Image Processing Using MATLAB*. Englewood Cliffs, NJ: Prentice-Hall, 2004.

[52] Y. S. Kim and W. Y. Kim, "Content-based trademark retrieval system using a visually salient feature," *Image Vis. Comput.*, vol. 16, no. 12/13, pp. 931–939, Aug. 1998.

[53] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *Proc. CVPR*, 2006, pp. 2161–2168.

[54] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. ICCV*, 2003, pp. 1470–1477.

[55] H. Stewénius and D. Nistér, Object recognition benchmark [Online]. Available: http://vis.uky.edu/%7Estewe/ukbench/

[56] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, Oct. 2004.

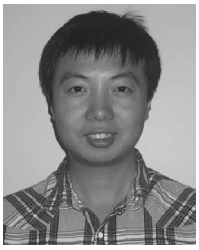
[57] D. Lowe, "Distinctive image features from scale-invariant key points," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[58] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments Univ. Massachusetts, Amherst, MA, Tech. Rep. 07- 49, 2007.

[59] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

Xiang Bai received the B.S., M.S., and Ph.D. degrees from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003, 2005, and 2009, respectively, all in electronics and information engineering.

He is currently an Associate Professor with the Department of Electronics and Information Engineering, HUST. From January 2006 to May 2007, he was with the Department of Computer Science and Information, Temple University, Philadelphia, PA. From October 2007 to October 2008, he was with the University of California, Los Angeles, as a joint Ph.D. student. His research interests include computer graphics, computer vision, and pattern recognition.



Bo Wang received the B.E. degree in electronic information engineering from Huazhong University of Technology and Science, Wuhan, China, in 2010. He is currently working toward the M.Sc. degree in the Department of Computer Science, University of Toronto, Toronto, ON, Canada.

His research interests include computer vision, machine learning, and numerical analysis.

Cong Yao received the B.S. degree in electronics and information engineering in 2008 from Huazhong University of Science and Technology (HUST), Wuhan, China, where he is currently working toward the Ph.D. degree in the Media and Communication Laboratory, Department of Electronics and Information Engineering.

His research has focused on computer vision and machine learning, particularly in the area of object detection and recognition in natural images.

Wenyu Liu received the B.S. degree in computer science from Tsinghua University, Beijing, China, in 1986 and the M.S. and Ph.D. degrees in electronics and information engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1991 and 2001, respectively.

He is currently a Professor and an Associate Dean of the Department of Electronics and Information Engineering, HUST. His current research areas include computer graphics, multimedia information processing, and computer vision.

Prof. Liu is a member of the IEEE System, Man, and Cybernetics Society.

Zhuowen Tu received the M.E. degree from Tsinghua University, Beijing, China, and the Ph.D. degree from Ohio State University, Columbus.

He is currently an Assistant Professor with the Laboratory of Neuro Imaging (LONI), Department of Neurology, with a joint appointment in the Department of Computer Science, University of California, Los Angeles (UCLA). He is also affiliated with the UCLA Bioengineering Interdepartmental Program, the UCLA Bioinformatics Program, and with Microsoft Research Asia, Beijing, China. Before joining LONI, he was a Member of Technical Staff with Siemens Corporate Research and a Postdoctoral Fellow with the Department of Statistics, UCLA.

Prof. Tu was a recipient of the David Marr Prize in 2003.