

# Local Geometry Adaptive Manifold Re-Ranking for Shape-Based 3D Object Retrieval

Ryutarou Ohbuchi  
University of Yamanashi  
4-3-11 Takeda, Kofu-shi  
Yamanashi, 400-8511, Japan  
ohbuchi@yamanashi.ac.jp

Yukinori Kurita  
University of Yamanashi  
4-3-11 Takeda, Kofu-shi  
Yamanashi, 400-8511, Japan  
g11mk014@yamanashi.ac.jp

## ABSTRACT

This paper proposes an improvement to Manifold Ranking algorithm used for search results ranking in the context of shape-based 3D model retrieval. Manifold Ranking algorithm by Zhou et al estimates, given a set of high-dimensional feature vectors, a lower-dimensional manifold on which the features lie. It then computes diffusion-based distances from a feature vector (or feature vectors) to the other feature vectors on the manifold. When applied to content-based retrieval, overall retrieval accuracy is significantly better than a “simple” fixed distance metric. However, in a small neighborhood of query, retrieval ranks obtained by a “simple” distance metric (e.g., L1-norm) performs better than those obtained by Manifold Ranking. Proposed re-ranking algorithm tries to combine ranking results due to both simple distance metric and Manifold Ranking in an automatic query expansion framework for better ranking results. Experimental evaluation has shown that the proposed method is effective in improving retrieval accuracy.

## Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Selection process.

## General Terms

Algorithms, Experimentation.

## Keywords

Content-based retrieval, re-ranking, automatic query expansion, manifold ranking, distance metric learning, 3D model retrieval.

## 1. INTRODUCTION

Three-dimensional (3D) shape models have been used in a variety of application areas, for example, mechanical and architectural design, entertainment, medical diagnosis, and for archaeology. Effective and efficient management methods for 3D models, especially via content based retrieval based on their shape, has become an important topic to be studied.

Generic processing pipeline for shape-based retrieval of 3D models starts with extraction of shape descriptors, or features, from query and database 3D models. Then, based on these features, database models are ranked based on their similarity to the query. In earlier 3D model retrieval algorithms, ranking has been performed by using a “fixed” distance, such as L1-norm.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM’12, October 29–November 2, 2012, Nara, Japan.

Copyright 2012 ACM 978-1-4503-1089-5/12/10...\$15.00.

However, learned distance metric adapted to the feature distribution is known to significantly improve retrieval accuracy. Such distance metric learning, whether it is supervised, unsupervised, or semi-supervised, has been *quite* popular in the field of text, image or video retrieval (e.g., [2]).

Manifold Ranking (MR) by Zhou et al [11, 12] is a powerful distance metric learning algorithm. It learns distribution of high-dimensional features in input feature space by connecting input features into a graph based on their proximity. Distances among feature points on the graph, which is an approximation of a low-dimensional manifold of features, are computed by using diffusion-like process on the graph. MR significantly improves retrieval accuracy (e.g., [2]).

While quite powerful, we found a weakness in MR. Our experience shows that MR often yields lower recall at top  $k$  retrievals, for example,  $k \lesssim 30$ , than a “simple” distance such as L1-norm. Figure 1 shows such a case in shape-based 3D model retrieval. It plots recall against retrieval rank for a simple distance metric L0.5-norm (broken line) and learned distance metric by MR algorithm (solid red line). For top ranked retrievals, e.g., top 50 or so, simple, fixed distance of L0.5-norm yields higher recall than MR. For lower ranked retrievals, e.g., rank > 100 or so, MR yields higher recall. Overall, MR has much higher accuracy. (In this example, its ground truth has 88 entries. Consequently, recall value for the “Ideal” retrieval algorithm saturates at 1.0 beyond rank 88.)

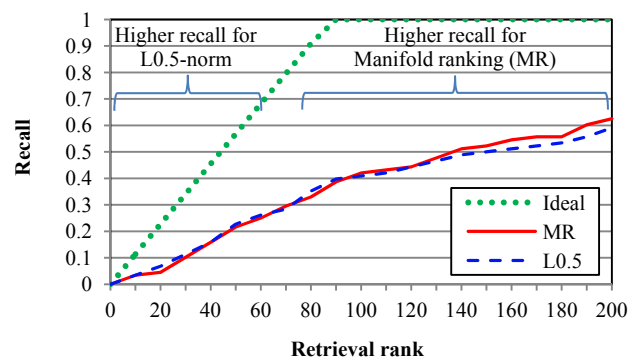


Figure 1. For top-ranked retrievals (e.g., rank < 50), simple distance (L0.5) produces better recall than Manifold Ranking (MR). Overall, however, MR produces significantly higher retrieval accuracy.

## 2. RELATED WORK

In the field of text, image, and video retrieval, distance metric learning has been embraced for some time. Distance metric learning has also been used in 3D model retrieval. Non-linear dimension reduction based on *manifold learning* has been employed [6]. A manifold learning algorithms in general

assumes and tries to find a low dimensional manifold on which data samples lie. For example, Locally Linear Embedding [8] creates a graph of input data samples based on proximity (e.g., by L1-norm) of the samples in their ambient, high dimensional feature space. The graph is an estimate of lower dimensional manifold on which the samples lie. By projecting original data samples onto an approximation of the manifold, lower dimensional representations of data samples are obtained.

Another popular manifold-based distance metric learning algorithm is *Manifold Ranking (MR)* by Zhou et al [11, 12, 2]. Similar to manifold learning, MR creates, as an estimate of low-dimensional data manifold, a graph connecting data samples by their proximity. Distances among the features are computed by diffusing relevance value on the graph, typically using an iterative process. MR has been shown to significantly improve overall retrieval accuracy for 3D model retrieval compared to a fixed “simple” distance such as L1-norm or Kullback-Leibler Divergence [7]. However, we discovered thorough experience that, for data samples that lie close to the query, i.e., higher ranked retrievals, a fixed, simple distance produces better retrieval results than MR. MR does outperform “simple” distances in terms of overall retrieval accuracy and retrieval accuracy for lower-ranked results. We suspect that the process of graph formation and relevance value diffusion may be erasing small geometrical details of the feature space near the query.

This paper proposes an algorithm that combines strength of both simple distance such as L1-norm and Manifold Ranking based learned distances for a better ranking results.

### 3. LOCAL GEOMETRY ADAPTIVE MANIFOLD RE-RANKING

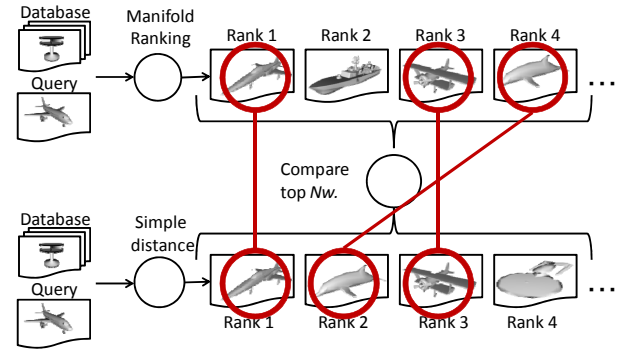
Proposed algorithm tries to combine higher accuracy of fixed distance near the query with overall strength of MR. The proposed algorithm is a simple 2-stage, automatic query expansion algorithm as described below;

- (1) **Initial Retrieval:** Two sets of ranked lists are retrieved by MR-based adaptive distance as well as by a simple distance such as L1-norm.
- (2) **Trusted retrieval set selection:** A set  $T$  of “trusted retrievals” are selected by comparing top ranked  $w$  retrievals, that is,  $w$ -nearest neighbors to the query, from the two ranked lists in (1). The parameter  $w$  is called *query expansion window size*. The set  $T$  consists of retrievals that appeared within top  $w$  of both ranked lists. (See Figure 2.) There are  $\|T\| = l \leq w$  members in  $T$ .
- (3) **Re-ranking:** MR-based ranking is performed by using the set  $T$  as the set of expanded queries.

#### 3.1 Ranking by Simple Distance

There are many different simple, or “non-data-adaptive” distances. To compute a distance among a pair of  $n$ -dimensional feature vectors  $x_i$  and  $x_j$ , we experimented with three different distances;  $Lp$ -norm  $d_{Lp}(x_i, x_j)$  for  $p=1.0$  and  $p=0.5$  and *Kullback-Leibler Divergence*  $d_{KLD}(x_i, x_j)$ . Due to space limitation, however, we will present experimental result for L1-norm,  $d_{L1}(x_i, x_j)$ , only.

$$d_{L1}(x_i, x_j) = \sum_{a=1}^n \|x_{ia} - x_{ja}\| \quad (1)$$



**Figure 2.** “Trusted retrieval set” is found by comparing ranked retrieval results produced by the MR and a simple distance. The set is used for second round of ranking by MR.

#### 3.2 Manifold Ranking and Re-Ranking

MR [11, 12] forms a graph  $G$  approximating a low-dimensional manifold based on proximity of feature points in an ambient  $n$ -dimensional feature space, i.e., by  $k$ -nearest neighbor connection. Each edge of the graph  $G$  is associated with an affinity value that controls diffusion of relevance through the edge. The affinity is computed as the affinity among a pair of features on the edge.

Let  $X = \{x_1, \dots, x_l, x_{l+1}, \dots, x_m\}$  be the set of  $n$ -dimensional features in  $\mathbb{R}^n$ . Vertices from  $x_1$  to  $x_l$  are the queries and the rest are the features points we want to rank. Let  $d : X \times X \rightarrow \mathbb{R}$  denote a distance metric on  $X$ , e.g., L1 norm or Cosine distance, that assigns a pair of points  $x_i$  and  $x_j$  a distance  $d(x_i, x_j)$ . Let  $f : X \rightarrow \mathbb{R}$  be a ranking function that assigns each  $x_i$  a ranking score  $f_i$ , forming a rank vector  $f = [f_1, \dots, f_m]^T$ , a  $m \times c$  matrix. Let the  $m$ -dimensional vector  $y = [y_1, \dots, y_m]^T$  be a label vector.

In the proposed algorithm, MR is used in two modes. For the initial retrieval step, MR computes the similarity ranks of 3D models in the database to a query. So we set  $y_i = 1$  for the query model only and  $y_i = 0$  for the rest, i.e., all the models in the database. For the re-ranking step, we give the set  $T$  having  $l$  features to the MR algorithm. Thus, we set  $y_i = 1$  for the features of the initial query and the features in the “trusted retrieval” set  $T$ , while  $y_i = 0$  is given to all the other features for the 3D models in the database.

Create an affinity matrix  $\mathbf{W}$  where  $\mathbf{W}_{ij}$  indicates the similarity between samples  $x_i$  and  $x_j$

$$\mathbf{W}_{ij} = \begin{cases} \exp(-d(x_i, x_j)^2 / 2\sigma^2) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The matrix  $\mathbf{W}$  is positive symmetric. Positive parameter  $\sigma$  defines the radius of influence of diffusion. Note that  $\mathbf{W}_{ii} = 0$  since there is no edge connecting a point with itself. As the distance  $d(x_i, x_j)$ , we use L1-norm  $d_{L1}(x_i, x_j)$  in (1). The matrix  $\mathbf{W}$  may be dense, but it could also be formed by connecting only  $k$  nearest neighbor of a feature in question to produce a sparse  $\mathbf{W}$ . Finite neighborhood defined by  $k$  affects geometry of the manifold embodied in  $\mathbf{W}$ , thus ranking accuracy, as we will see in experiments. The parameter  $k$  also affects computational cost of MR as  $k$  changes sparsity of  $\mathbf{W}$ .

From  $\mathbf{W}$ , a normalized *graph Laplacian*  $\mathbf{L}$  is formed;

$$\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2} \quad (3)$$

where  $\mathbf{D}$  is a diagonal matrix in which  $\mathbf{D}_{ij}$  equals to the sum of the  $i$ -th row of  $\mathbf{W}$ , that is,  $\mathbf{D}_{ij} = \sum_j \mathbf{W}_{ij}$ . The ranking vector  $\mathbf{F} = [f_1, \dots, f_n]^T$  that indicates the likelihood of vertices having a label can then be estimated by iterating the following equation until convergence;

$$f^{(t+1)} = \frac{1}{1+\mu}(\mathbf{I} - \mathbf{L})f^{(t)} + \frac{1}{1+\mu}\mathbf{Y} \quad (4)$$

The parameter  $\mu > 0$  is a regularization parameter, and affects retrieval performance and the convergence of the iteration above. Let  $f^*$  be the limit of the above iteration. Rank each point  $x_i$  as a label  $y_i = \arg \max_{j \leq c} f_{ij}^*$ . In the case above,  $f^*$  has a closed form solution;

$$f^* = \left( \mathbf{I} + \frac{1}{\mu} \mathbf{L} \right)^{-1} \mathbf{Y} \quad (5)$$

Computational cost is an issue to be considered if MR is to be used in an interactive segmentation loop. The cost of MR is dominated by the cost of meshing and computing  $f^*$ . The cost increases with the size of the matrix  $\mathbf{L}$ , that is, the number of samples  $m$ . For example, the matrix  $\mathbf{L}$  takes  $O(m^2)$  (if it is dense) to store and roughly  $O(m^3)$  to perform label propagation via matrix inversion.

## 4. EXPERIMENTS AND RESULTS

We performed experiments to evaluate effectiveness of the proposed re-ranking algorithm. In Section 4.1, we first evaluate effect of MR neighborhood size  $k$  on retrieval accuracy. We then evaluate the effect of the size of locality  $w$ .

Retrieval experiments are performed by using two benchmark databases: *McGill Shape Benchmark* (MSB) [9] for highly articulated (non-rigid) but less geometrically detailed shapes, and *SHape REtrieval Contest* (SHREC) 2006 benchmark [10] for a set of quite diverse, rigid, and relatively detailed shapes. MSB contains models represented as densely sampled closed manifold mesh, while SHREC 2006 contains 3D models represented by using polygons soup, meshes having high variance in sampling density, etc. MSB uses models in the dataset as query for evaluation, while SHREC 2006 uses a set of 30 out-of-sample queries for evaluation. As the performance index, we use Mean Average Precision (MAP) in the following.

We used a 3D model feature vector produced by *Bag-of-Features Dense SIFT* (BF-DSIFT) [1] algorithm. BF-DSIFT is a view-based 3D shape comparison algorithm. Given a 3D model, BF-DSIFT renders range images from multiple viewpoints located uniformly in the solid angle. From each range image, SIFT features are extracted at densely and randomly placed feature points. Having multiple viewpoints and use of (in-plane) rotation invariant SIFT feature provides 3D rotation invariance. A set of over 10k SIFT features are extracted from a 3D model, which is then integrated into a feature vector having 30k or so dimension by using bag-of-features approach. Combination of SIFT, a multi-scale local visual feature, and bag-of-features integration enables the BF-DSIFT to have invariance against articulation and global deformation of 3D models. BF-DSIFT is one of best-performing algorithms; it placed 1st in SHREC 2012 Generic 3D model track [3].

### 4.1 MR meshing neighborhood size $k$

We evaluated the impact of MR neighborhood size  $k$  on retrieval accuracy. A small  $k$  might capture local geometry of feature

manifold better for higher retrieval accuracy. A small  $k$  also produces a sparser matrix, which leads to less memory footprint and less computational cost. On the other hand, if  $k$  is too small, the graph as an estimate of feature manifold would break up, resulting in lower retrieval accuracy.

Figure 3 and Figure 4 show plots of MAP against neighborhood size  $k$  and window size  $w$  for MSB and SHREC 2006 benchmarks. These plots used L1-norm for MR meshing and BF-DSIFT as input feature. For  $w=20$ , which is the best performing windows size for both MSB and SHREC 2006, MAP scores peak at  $k=5$  for MSB and  $k=20$  for SHREC 2006. For the values of  $k$  beyond these peaks, performance degradation is gradual and small ( $<0.05$ ). For the values of  $k$  smaller than these peaks, i.e.,  $k < 5$ , retrieval accuracies drop steeply.

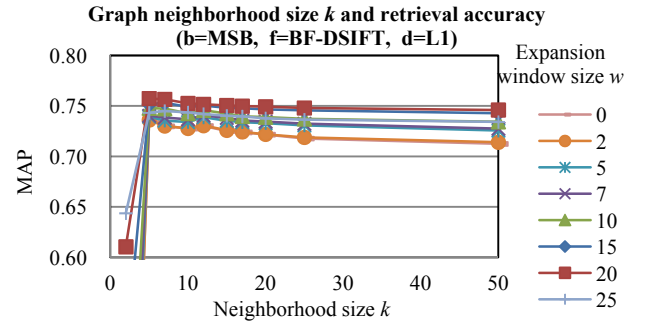


Figure 3. Neighborhood size  $k$  for MR meshing and retrieval accuracy using BF-DSIFT feature measured on MSB benchmark (L1 norm). MAP peaks at  $k=5$ .

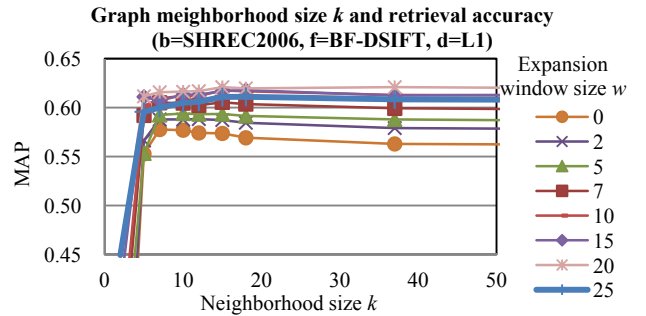


Figure 4. Neighborhood size  $k$  for MR meshing and retrieval accuracy using BF-DSIFT feature measured on SHREC2006 benchmark (L1 norm). MAP peaks at  $k=10$ .

### 4.2 Query expansion window size $w$

We evaluated the impact of query expansion window size  $w$  on retrieval accuracy. Figure 5 and Figure 6 show, respectively, results for MSB and SHREC 2006 benchmarks. Both used L1-norm for MR meshing and BF-DSIFT as input feature. In each plot, query expansion window size  $w=0$  represents the baseline case, i.e., MR without proposed re-ranking.

For both benchmark databases, retrieval accuracy measured by using MAP improved significantly and consistently for non-zero  $w$ , i.e., if proposed re-ranking is used. As the figures show, MAP values plotted against  $w$  have peak at around  $w=20$ . Shape of the plots and the positions of the peaks are similar among the two benchmarks. Also, MR neighborhood size  $k$  does have only small impact on the position of peaks, except for the cases where  $k=2$ . Table 1 summarizes best MAP scores for both baseline and proposed algorithms.

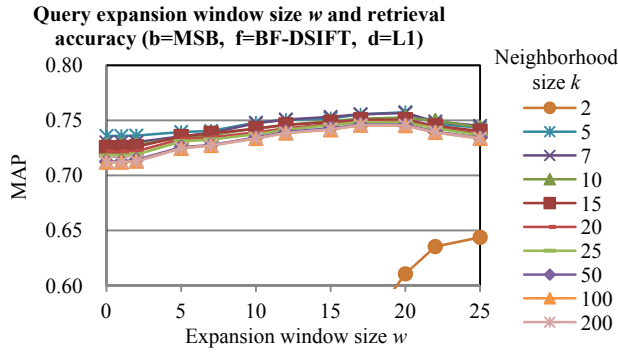


Figure 5. Query expansion window size  $w$ , graph neighborhood size  $k$ , and retrieval accuracy (MAP) measured using BF-DSIFT feature on MSB benchmark (L1 norm). ( $w=0$  is the baseline for respective values of  $k$ .)

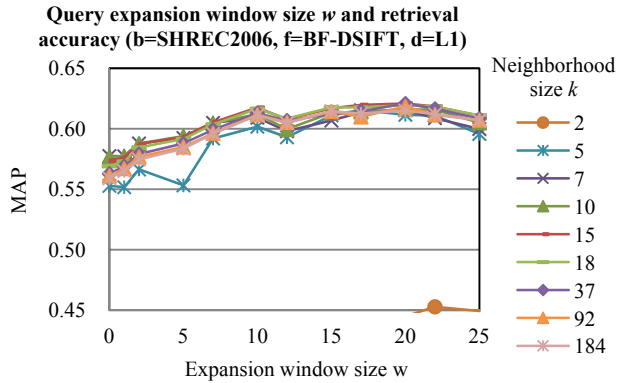


Figure 6. Query expansion window size  $w$ , graph neighborhood size  $k$ , and retrieval accuracy (MAP) measured using BF-DSIFT feature on SHREC 2006 benchmark (L1 norm). ( $w=0$  is the baseline for respective values of  $k$ .)

Table 1. Best MAP scores for original MR and proposed re-ranking algorithms.

	MR (i.e., $w=0$ )	Proposed re-ranking	MAP increase
MSB	0.7358 ( $k=5$ , $w=0$ )	0.7574 ( $k=5$ , $w=20$ )	+0.0216
SHREC 2006	0.5830 ( $k=7$ , $w=0$ )	0.6256 ( $k=18$ , $w=15$ )	+0.0426

## 5. SUMMARY AND CONCLUSION

Distance metric learning using manifold assumption is known to provide significant improvement in ranking performance for image, text and other retrieval tasks. One such algorithm is Manifold Ranking (MR) algorithm by Zhou, et al [11, 12]. Experiments show that MR algorithm provides significant improvement in overall retrieval accuracy compared to a simple, fixed distance metric such as L1-norm. Ranking performance of MR is less than a simple, fixed distance, however, at a small proximity of query, i.e., for high-rank retrievals. This paper proposed a novel re-ranking algorithm that takes advantage of both simple, fixed distance and MR-based learned distance.

The proposed algorithm is an automatic query expansion algorithm. The algorithm first performs retrieval by using both simple, fixed distance and MR to have two ranked lists of

retrievals. Then, by comparing the top  $w$  results from the two lists, a set of “trusted retrievals” are selected. The union of the trusted retrieval set and the original query becomes an expanded set of queries to be fed to MR algorithm for a re-ranked retrieval results. Experimental evaluation using multiple benchmark databases showed that the proposed algorithm is effective in improving retrieval accuracy. For example, performance measured in Mean Average Precision (MAP) increased by 7% from 0.5830 to 0.6256 in the case of SHREC 2006 benchmark. A drawback of the proposed algorithm is an increased cost of ranking, which involves two MR rankings and a ranking using simple distance. We plan to investigate a method to alleviate this drawback.

## ACKNOWLEDGMENTS

This research has been funded in part by the Ministry of Education, Culture, Sports, Sciences, and Technology of Japan (No. 18300068).

## REFERENCES

- [1] T. Furuya, R. Ohbuchi, Dense Sampling and Fast Encoding for 3D Model Retrieval Using Bag-of-Visual Features, *Proc. ACM Intern'l Conf. on Image and Video Retrieval 2009 (CIVR 2009)*, July 8-10, (2009)
- [2] J. He, M. Li, H-J. Zhang, H. Tong, C. Zhang, Manifold-ranking based image retrieval, *Proc. ACM Multimedia 2004*, 9-16, 2004.
- [3] B. Li, A. Godil, et al, SHREC'12 Track: Generic 3D Shape Retrieval, to appear, *Proc. Eurographics Workshop on 3D Object Retrieval*, 2012
- [4] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV*, **60**(2), 91-110, (2004).
- [5] McGill 3D Shape Benchmark, <<http://www.cim.mcgill.ca/~shape/benchMark/>>
- [6] R. Ohbuchi, A. Yamamoto, J. Kobayashi, Learning semantic categories for 3D Model Retrieval, *Proc. ACM Workshop on Multimedia Information Retrieval 2007*, pp. 31-40 (2007).
- [7] R. Ohbuchi, T. Furuya, Distance Metric Learning and Feature Combination for Shape-Based 3D Model Retrieval, poster paper, *Proc. ACM workshop on 3D object retrieval 2010*, Firenze, Italy, (2010)
- [8] S.T. Roweis, L.K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, **290**(5500), pp. 2323-2326 (2000).
- [9] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, S. Dickinson, Retrieving Articulated 3D Models Using Medial Surfaces, *Machine Vision and Applications*, **19**(4), 261-275, 2008.  
<http://www.cim.mcgill.ca/~shape/benchMark/>.
- [10] R. C. Velkamp, SHREC2006 3D Shape Retrieval Contest, Utrecht University Dept. Information and Computing Sciences, Tech. Report UU-CS-2006-030 (ISSN: 0924-3275).
- [11] D. Zhou, O. Bousquet, T.N. Lal, J. Weston and B. Schölkopf, Learning with Local and Global Consistency, *Proc. NIPS 2003*, 2003.
- [12] D. Zhou, J. Weston, A. Gretton, O. Busquet, and B. Schölkopf, Ranking on Data Manifolds, in *Proc. NIPS 2003*, 2003.