

Distance-Free Image Retrieval Based on Stochastic Diffusion over Bipartite Graphs

Christian Bauckhage
Deutsche Telekom Laboratories
10587 Berlin, Germany
<http://www.telekom.de/laboratories>

Abstract

We propose an approach to image retrieval that does not require any distance computations. The idea is to represent images and corresponding image features by means of the two sets of vertices of a bipartite graph. Even though in such a graph the images are not directly related, the degrees to which the features are present in an image allow for defining partial orders. If the degrees of presence are normalized such that they form probability distributions, similarity rankings result from the stationary distributions of stochastic diffusion processes over the graph. The method is closely related to recent approaches to ranking on manifolds but does not involve the computation of parameterized affinity and Laplacian matrices. Experiments with a standard image retrieval data set demonstrate the efficacy of the approach. Compared to a corresponding distance-based approach, it yields a higher overall precision.

1 Introduction

Content-based image retrieval (CBIR) from large databases has become a task of considerable practical importance. Admen, artists, designers, and journalists need fast access to appropriate icons or pictures to illustrate advertisements, journals, jingles or whatever else requires visual amelioration nowadays. However, the sheer size and speed of growth of present day image repositories create a crucial problem: consistent semantic annotations can hardly be provided single-handedly anymore. Neither can teamwork guarantee consistency. Experience with *folksonomies* gathered and maintained by online communities shows that spurious and ambiguous labels occur inevitably. Figure 1 illustrates what this implies in practice; it displays a choice from the 40 top ranking results obtained from typing “tiger” into Google’s image search.

State of the art retrieval systems therefore apply computer vision techniques that are fine tuned to the task at hand by means of user feedback [3, 12, 14]. In the so called *human-in-the-loop* approach, the user repeatedly rates selections of images according to how well they match the current query. Based on this relevance feedback, characteristics of appropriate and inappropriate images are determined and a hopefully better suited set of images is retrieved from the database. This interactive process continues until the user’s demands are met.

In a series of influential papers, Rui and Huang [11, 12] presented interactive CBIR systems based on a hierarchical model that combines different features and adaptable



Figure 1: Some of the top ranking results when searching Google Images for “tiger”.

distance measures. Even though their model and many of its successors enable flexible searches for images similar to the user’s intent, the way the different distances between features are defined appears to be solely technically motivated and is hard to grasp intuitively. More recent approaches [3, 14] apply more sophisticated reasoning and adaptation processes, but at their heart, too, lies the problem of defining distances between images that would allow for producing similarity rankings.

The reason why we emphasize this issue here is that it became clear some time ago that sets of images of a semantic class tend to form nonlinear manifolds whose global structure cannot be captured by simple metrics (see the examples in [2, 10, 13]).

Dealing with the problem of CBIR, the question then is how define similarities between objects residing on such manifolds. Or, in other words, what is needed is a method to rank such objects. As a matter of fact, this problem has been addressed in several recent contributions [1, 7, 8, 15, 16]. It has even been studied with respect to information retrieval in general [5] and image retrieval in particular [6]. Since these approaches are closely related to the idea presented in this paper, we will discuss them in more depth later on. For now, we simply point out that all these approaches derive the global structure of a set of data by considering local relations among individual elements which are again based on some notion of distance.

In this paper, we consider only a single iteration in an interactive CBIR system and focus on the problem of image ranking. Our approach determines similarities among images but does not require any distance computations. The idea is to represent a collection of images and a set of meaningful image features by means of the two sets of vertices of a bipartite graph. Assuming the edges between images and features to denote transitions in a Markov process immediately provides an ordering scheme: if we model a user query as an initial distribution over the vertices corresponding to images, a ranking results from the stationary distribution of a corresponding Markov chain that starts from this initial state.

In the next section, we detail this idea and the computational approach. We will see that there is a simple closed form solution to compute image rankings from an arbitrary query. We will discuss that, similar to the approaches in [1, 7, 8, 15, 16], our approach leads to a graph diffusion kernel. In contrast to existing methods, however, the kernel naturally results from the probabilistic model and its derivation does not require manual adjustment of free parameters. In section 3, we present experiments that demonstrate the efficacy of the proposed approach. On a standard data set it yields useful precision and outperforms a distance-based retrieval method considered for baseline comparison. Finally, section 4 concludes this paper and points out promising next steps of research.

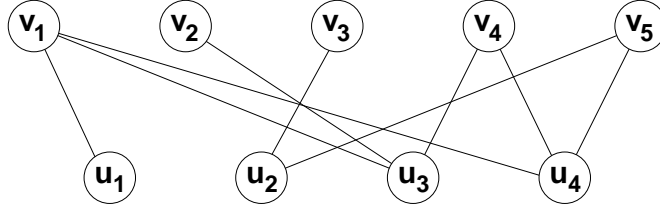


Figure 2: Example of a bipartite graph. Although there are no direct relations among the vertices u_i , their relations with the vertices v_j define a similarity and thus allow for partial ordering. With respect to vertex u_3 , for instance, the order is $u_3 \supseteq u_4 \supseteq u_1 \supseteq u_2$

2 Ranking as a Markov Process over Bipartite Graphs

The idea for the CBIR approach presented in this paper occurred while we were exploring novel mechanisms for collaborative filtering for automatic recommender systems. In the discussion that follows, we will thus frequently resort to rather metaphorical language and make use of terms such as *vote for* or *rate* which we feel convey the underlying ideas.

2.1 Mathematical Model

Assume labeled bipartite graph $G = (V, E)$ as shown in Fig. 2. Its sets of vertices V is partitioned such that $V = V_1 \cup V_2$ and $V_1 \cap V_2 = \emptyset$. The n vertices u_1, u_2, \dots, u_n in the set V_1 correspond to entities (such as users, images, ...). In a slight abuse of notation we will identify vertices and their labels and represent a labeling of the vertices in V_1 by means of a vector $\mathbf{u} = [u_1, u_2, \dots, u_n]^T$. The m vertices v_1, v_2, \dots, v_m in the set V_2 correspond to rated items or features that are voted for (e.g. books, RGB color bins, gradient directions, ...) and their labels are stored in a vector $\mathbf{v} = [v_1, v_2, \dots, v_m]^T$.

In a recommender systems, each entity $u_i \in V_1$ votes for (a subset of) the items in V_2 . Dealing with CBIR, we may think of the votes as indicators to what extend a certain feature in V_2 is present in an image represented by u_i . In both cases, votes or frequency counts can be represented by means of directed, weighted edges (see Fig. 3(a)).

Even though there are no immediate relations (i.e. no edges) among the elements in V_1 , their voting behavior allows for determining partial orders. Given an entity u_i , its fellow entities can be ranked according to how much their voting behavior resembles the one of u_i . In contrast to common distance measures between vectors of votes or frequency counts, the bipartite graph model seamlessly accounts for indirect relations as well. In the example shown in Fig. 2, for instance, u_2 is related to u_3 and u_4 alike. However, while the nature of its relation to u_4 is of first degree because both entities share a vote, its relation to u_3 is a second degree relation because it is mediated through u_4 .

The key idea is now to understand relations of arbitrary degree as the outcome of a stochastic diffusion process over the bipartite graph. To this end, we normalize the votes cast by an entity so that they sum to 1. If all the votes of all the entities are stored in a column stochastic $m \times n$ matrix \mathbf{R} , and entity vectors are normalized so that they sum to 1, too, individual or weighted combined ratings result from $\mathbf{v}_t = \mathbf{R}\mathbf{u}_t$. With these assumptions, we obviously are considering probabilistic mappings from V_1 to V_2 .

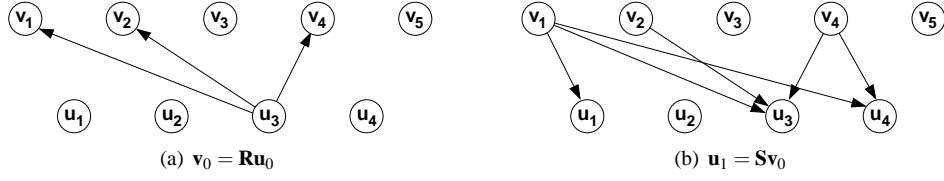


Figure 3: Example of the beginning of a stochastic diffusion process over a bipartite graph. Starting with the distribution $\mathbf{u}_0 = [0010]^T$ produces a distribution \mathbf{v}_0 which in turn leads to the updated distribution \mathbf{u}_1 .

Given the transition matrix \mathbf{R} , each item $v_j \in V_2$ can deduce which entities do vote for it for this information is essentially contained in the transpose of the transition matrix. If $\mathbf{S} \propto \mathbf{R}^T$ was normalized so that it is a column stochastic matrix, too, a set of rated items can (in turn) vote for entities (see Fig. 3(b)). An updated distribution over the entities in V_1 would then result from

$$\mathbf{u}_{t+1} = \mathbf{S}\mathbf{v}_t = \mathbf{S}\mathbf{R}\mathbf{u}_t \stackrel{!}{=} \mathbf{H}\mathbf{u}_t. \quad (1)$$

Note that the $n \times n$ matrix \mathbf{H} introduced in last step of this derivation is a doubly stochastic matrix whose rows columns and rows sum to 1. It is square and non-negative and its eigenvalues λ_k are characterized by $|\lambda_k| \leq 1$.

Also, note that \mathbf{H} defines a Markov process over the set V_1 . Therefore, even though no direct relation among the $u_i \in V_1$ were available in the first place, we now have a tool for ranking. Assume an initial distribution \mathbf{u}_0 with only a few non zero entries. Then, after t steps, the probabilities in $\mathbf{u}_t = \mathbf{H}^t \mathbf{u}_0$ will be higher for entities which are more closely related to the initially active elements and less high for less closely related ones.

However, in this most simple form, the model cannot produce reasonable rankings if the underlying Markov chain is irreducible and contains positive-recurrent states. In this (practically very likely) case, the process converges to a uniform distribution over the elements in V_1 which does not allow for any ranking. We therefore assume the initial distribution \mathbf{u}_0 to be a steady source of probability mass that constantly feeds the stochastic process. With this modification, the update rule for distributions is given by

$$\mathbf{u}_{t+1} = \frac{1}{2} \left[\mathbf{H}\mathbf{u}_t + \mathbf{u}_0 \right] \quad (2)$$

where the scaling factor $\frac{1}{2}$ ensures that \mathbf{u}_{t+1} does sum to 1 just as \mathbf{u}_t and \mathbf{u}_0 do. With some algebra it is easy to see that, written as a power series, the recursive expression in (2) amounts to

$$\mathbf{u}_t = \left(\frac{1}{2} \mathbf{H} \right)^t \mathbf{u}_0 + \frac{1}{2} \sum_{i=0}^{t-1} \left(\frac{1}{2} \mathbf{H} \right)^i \mathbf{u}_0 \quad (3)$$

Recall that \mathbf{H} is a doubly stochastic matrix whose eigenvalues λ_k satisfy $|\lambda_k| \leq 1$. For the limit $t \rightarrow \infty$, we therefore have

$$\lim_{t \rightarrow \infty} \left(\frac{1}{2} \mathbf{H} \right)^t = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \sum_{i=0}^{t-1} \left(\frac{1}{2} \mathbf{H} \right)^i = \left[\mathbf{1} - \frac{1}{2} \mathbf{H} \right]^{-1}. \quad (4)$$

Hence, the iteration in equation (2) is guaranteed to converge. Once the process has converged, the vector \mathbf{u} it converged to is characterized by $\mathbf{u} = \frac{1}{2} [\mathbf{H}\mathbf{u} + \mathbf{u}_0]$ which directly leads to the closed form solution

$$\mathbf{u} = \frac{1}{2} \left[\mathbf{1} - \frac{1}{2} \mathbf{H} \right]^{-1} \mathbf{u}_0. \quad (5)$$

Therefore, given an arbitrary initial distribution \mathbf{u}_0 that might represent a single entity or –just as well– a mixture of entities, we can immediately determine the corresponding stationary distribution and the ranking it implies.

2.2 Discussion

It is interesting to note that the matrix in equation (5) constitutes a diffusion kernel [8]. In fact, from the derivation, we recognize another instance of the *kernel trick*. The similarities among vectors $\mathbf{u} \in \mathbb{R}^n$ that are contained in $\mathbf{H} = \mathbf{S}\mathbf{R}$ result from mapping the vectors back and forth to a (usually higher dimensional) space \mathbb{R}^m .

Diffusion kernels for the purpose of computing similarities on manifolds or graphs have recently been studied by several authors [1, 7, 8, 15, 16]. In two contributions closely related to this paper, Zhou et al. [15, 16], investigate the problem of ranking on manifolds. The manifolds they are concerned with are represented by means of adjacency graphs. Given an unstructured set of feature vectors, they compute a matrix that represents local structures in the data by means of the distances between each data point and its k nearest neighbors. The adjacency matrix is then transformed into a similarity matrix \mathbf{K} using a Gaussian kernel with parameter σ . Given \mathbf{K} , they show that diffusion processes on this adjacency graph are governed by the matrix $(1 - \alpha) [\mathbf{1} - \alpha \mathbf{K}]^{-1}$. This, of course, closely resembles the result in (5).

In fact, from setting $\alpha = \frac{1}{2}$, we recognize stochastic diffusion over a bipartite graph to be a special case of the problem studied [15, 16]. However, some comments appear to be in order. While our derivation did not involve any free parameters, the approach by Zhou et al. requires at least three of them (k , σ , α). Moreover, while our approach avoids the computation of distances between vectors of ratings or features, the approach by Zhou et al. requires distance computation for constructing the adjacency matrix as well as the corresponding similarity matrix. Finally, the matrix \mathbf{H} in our approach is a stochastic matrix and thus allows for a concise interpretation of the ranking procedure in terms of a Markov process. The matrix \mathbf{K} in the approach by Zhou et al., in contrast, eludes such an interpretation.

Ranking on manifolds has already been applied in systems for document and image retrieval [5, 6]. However, to the best of our knowledge, all known such systems consider diffusion processes over adjacency graphs that represent local neighborhoods similar to the way discussed above. They therefore leave the user with the problem of choosing suitable distances and parameters. Our approach, on the other hand, is parameter-free. In the next section, we present initial experiments which demonstrate that it nevertheless yields useful results for the problem of CBIR.

3 Experiments

In this section, we report first results obtained from our approach to image similarity ranking. Note that, in our experiments, we did not pay too much attention to the selection of features suitable for the task of CBIR. Therefore, the figures and examples presented below should not be considered the maximum achievable performance. Rather, they are meant to illustrate the potential of ranking based on diffusion over bipartite graphs.

3.1 Setting

All our experiments considered the Corel 1000 data set [9]. It contains 1000 color images showing scenes or objects from 10 different categories; for each category, there are a 100 examples.

Since the idea of the *degree of presence* of a feature, which we alluded to in the last section, naturally translates to the use of histograms, we considered histogram-based descriptors to characterize entire images in the data set. In order to represent information due to the geometric structure of the image content, we decided to apply histograms of oriented gradients as introduced by Dalal and Triggs [4]. We used 12 bins to store gradient directions computed over a 9×9 grid of cells. The nonlinear normalization of different histograms was computed with respect to 3×3 blocks of cells. In order to represent information contained in the color distributions of the images, we adopted the idea by Dalal and Triggs to color histograms. Here, we considered 5×5 cells which again were normalized using 3×3 blocks. The color histograms in each of the cells contained 20 bins; the corresponding prototypical colors were determined from clustering the pixels of all images in the database into different sets. Other than that, no preprocessing steps were applied; in particular, we did not perform brightness adjustments or color normalization such as proposed in [4].

Given these image descriptors, we tested how our approach performed when the descriptors were considered individually as well as how it performed when they were combined into a larger vector. For baseline comparison, we also verified how a retrieval procedure performed that determines image similarities based on the cosine distance between feature vectors.

The figures in the Tables 1 to 3 resulted from issuing 10 different queries for each category and averaging over the results. In accordance with the traditional approach in information retrieval, we characterize the different algorithms in our test with respect to the *precision* they achieved.

3.2 Results

Tables 1, 2, and 3 list the precision values *at 5*, *at 10*, and *at 20*, respectively, and thus indicate how many relevant documents were returned among the top 5, top 10, and top 20 ranking documents. Results obtained from the histogram of gradients features are found in the columns marked *HOG*, the ones obtained from histograms of colors are displayed in the columns marked *HOC*, results yielded by the combined descriptors are labeled *both*.

Although some images seem to defy retrieval (e.g. the pictures of Mountains), the results obtained from stochastic diffusion processes over bipartite graphs generally appear reasonable and useful. Moreover, on average, our approach consistently outperforms the

	stochastic diffusion			cosine distance		
	HOG	HOC	both	HOG	HOC	both
New Guinea	82	82	82	10	50	44
Beaches	80	90	94	46	54	44
Rome	40	20	32	14	44	32
Busses	80	52	78	76	56	78
Dinosaurs	96	100	100	72	100	100
Elephants	20	78	70	22	66	50
Flowers	60	72	78	34	94	54
Horses	22	76	74	96	90	94
Mountains	4	24	18	8	42	34
Food	26	36	66	0	58	58
average	51	63	69	38	65	59

Table 1: Precision @ 5 obtained on the Corel 1000 data set.

	stochastic diffusion			cosine distance		
	HOG	HOC	both	HOG	HOC	both
New Guinea	76	76	76	5	52	41
Beaches	73	84	90	40	39	41
Rome	35	21	33	14	37	31
Busses	74	45	74	71	48	71
Dinosaurs	89	97	97	59	100	100
Elephants	21	68	65	15	61	42
Flowers	62	74	78	33	88	49
Horses	19	68	64	92	85	89
Mountains	4	21	17	5	36	25
Food	22	32	59	5	55	48
average	48	59	65	34	60	54

Table 2: Precision @ 10 obtained on the Corel 1000 data set.

baseline method, if it considers the combination of gradient and color features. Preliminary results like this are promising and justify further work on CBIR based on parameter-free diffusion over bipartite graphs.

Figures. 4 through 6 exemplify another interesting and promising feature of our approach: since it avoids the computation of distances, it does not only apply to ranking with respect to individual elements on a manifold but can be seamlessly applied in order to rank with respect to sets of elements. The figures illustrate, how this can aid CBIR.

In its lower row, Fig. 4 shows the top 5 ranking images that were returned when the image in the upper row was used as the query example. The ranking resulted from using the combined gradient and color features and starting the Markov chain with an initial

	stochastic diffusion			cosine distance		
	HOG	HOC	both	HOG	HOC	both
New Guinea	65	65	65	9	54	37
Beaches	69	77	82	32	38	40
Rome	30	18	26	15	27	26
Busses	66	37	71	60	48	64
Dinosaurs	80	90	90	42	100	98
Elephants	20	56	53	16	57	42
Flowers	58	66	74	27	80	50
Horses	21	58	60	88	79	86
Mountains	6	22	17	10	35	25
Food	19	34	52	7	52	40
average	43	52	59	30	57	51

Table 3: Precision @ 20 obtained on the Corel 1000 data set.

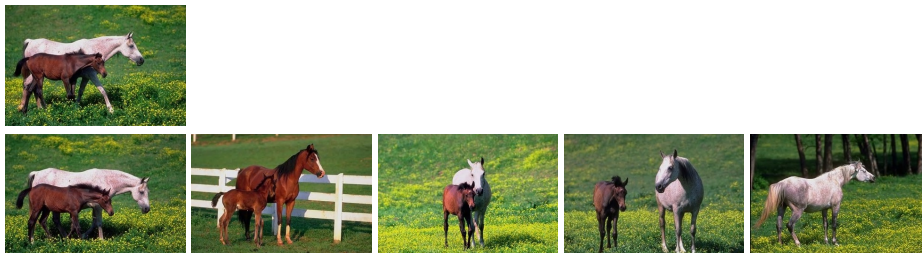


Figure 4: A single query image and the 5 top ranking results.

distribution $\mathbf{u}_0 = [0 \dots 010 \dots 0]^T$. Figures 5 and 6 show the outcome of the process when started with a distribution $\mathbf{u}_0 = \frac{1}{M}[0 \dots 010 \dots 010 \dots 0]^T$ where $M = 3$ elements were set to $\frac{1}{M}$. From Fig. 5 we see that, if these elements index visually similar images, the retrieved images appear similar to these images, too. If the initial distribution covers a set of less similar images, the ones that will be returned among the top ranking images will also show a greater variety (see Fig. 6).

4 Summery and Outlook

In this paper, we described a novel approach to image ranking for content-based image retrieval. The interesting characteristics of this approach are that it is parameter-free and that it determines image similarities without computing distances. Given a collection of images together with a corresponding set of normalized feature vectors, the idea is to understand both sets as the disjoint sets of vertices of a bipartite graph. If the edges between images and features are assumed to denote transitions in a Markov process and if given queries are taken to be the initial distribution, an ordering with respect to a query results

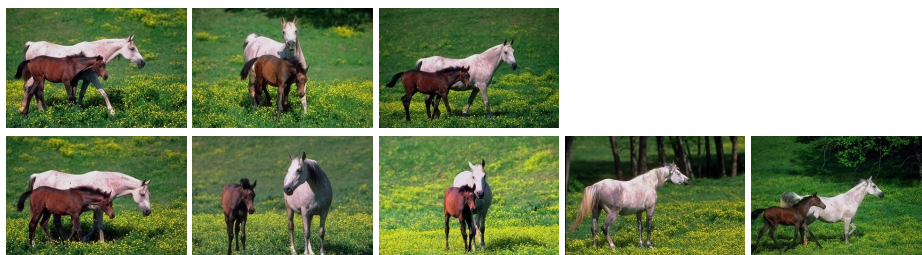


Figure 5: Three similar query images and the 5 top ranking results.

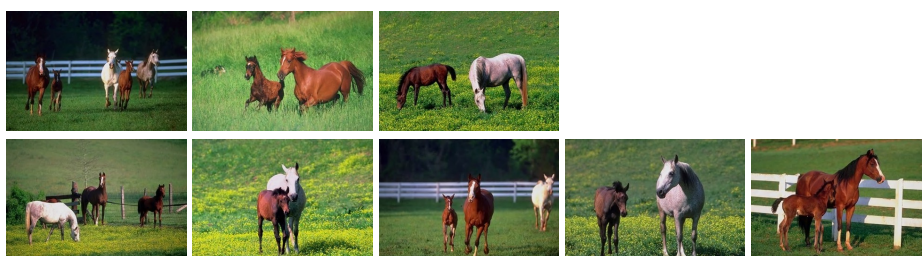


Figure 6: Three less similar query images and the 5 top ranking results.

from the stationary state of the chain. By design –and in contrast to other recent approaches to manifold ranking– our approach allows for a rigorous interpretation in terms of Markov processes. Since these are completely characterized by the underlying stochastic matrix, a user does not have to adjust free parameters and distance measures. On the contrary, feature frequency counts or histograms immediately lead to necessary transition probabilities.

Preliminary results obtained with this approach are promising and justify further investigation as to what features might further improve precision. In addition, the method itself offers interesting perspectives for future research. An obvious idea is to apply it to classification: given a feature vector \mathbf{v} derived from an unknown input image and a set of known images, the new image can be classified by, for instance, a majority count of the top ranking entities in the vector \mathbf{u} that results from a query with the initial distribution $\mathbf{u}_0 = \mathbf{S}\mathbf{v}$. Another direction worth pursuing further appears from noting that equation (2) resembles the systems one deals with in linear quadratic control. The noticeable difference is that, in equation (2), the control matrix is set to $\mathbf{1}$. Especially from the point of view of interactive content-based retrieval, ways of adapting this matrix to better meet the user’s intent seem a worthwhile topic.

References

- [1] S. Agarwal. Ranking on Graph Data. In *Proc. ICML*, pages 25–32, 2006.

- [2] M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, 2003.
- [3] I.J. Cox, M.L. Miller, T.P.Minka, T. Papatomas, and P.N. Yianilos. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments. *IEEE Trans. on Image Processing*, 9(1):20–37, 2000.
- [4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *Proc. CVPR*, volume 2, pages 886–893, 2005.
- [5] F. Fous, A. Pirotte, J.-M. Renders, and M. Saerens. Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. *IEEE Trans. on Knowledge and Data Engineering*, 19(3):355–369, 2007.
- [6] J. He, M. Li, H.-J. Zhang, H. Tong, and C. Zhang. Manifold-Ranking Based Image Retrieval. In *Proc. ACM Int. Conf. on Multimedia*, pages 9–16, 2004.
- [7] H. Kashima, K. Tsuda, and A. Inokuchi. Kernels for graphs. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel Methods in Computational Biology*, pages 155–170. MIT Press, 2004.
- [8] R.I. Kondor and J.D. Lafferty. Diffusion Kernels on Graphs and Other Discrete Input Spaces. In *Proc. ICML*, pages 315–322, 2002.
- [9] J. Li and J.Z. Wang. Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088, 2003.
- [10] S. Roweiss and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.
- [11] Y. Rui and T. Huang. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):644–665, 1998.
- [12] Y. Rui, T. Huang, and S. Chang. Image Retrieval: Current Techniques, Promising Directions and Open Issues. *J. of Visual Communication and Image Representation*, 10(4):39–62, 1999.
- [13] J.F. Tenenbaum, V. de Silva, and J.C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(550):2319–2323, 2000.
- [14] J. Vogel and B. Schiele. Performance Evaluation and Optimization for Content-Based Image Retrieval. *Pattern Recognition*, 39(5):897–909, 2006.
- [15] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Schölkopf. Learning with Local and Global Consistency. In *Proc. NIPS*, number 16, pages 321–328, 2004.
- [16] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on Data Manifolds. In *Proc. NIPS*, number 16, pages 169–176, 2004.