ANISOTROPIC MANIFOLD RANKING FOR VIDEO ANNOTATION

Jinhui Tang [†], Xian-Sheng Hua [‡], Guo-Jun Qi [†], Tao Mei [‡], Xiuqing Wu [†]

[†] MOE-Microsoft Key Laboratory of Multimedia Computing and Communication,
University of Science and Technology of China, Hefei, 230027 China
[‡] Microsoft Research Asia, Beijing, 100080 China

ABSTRACT

Graph-based semi-supervised learning (SSL) has attracted lots of interests in machine learning community as well as many application areas including video annotation recently. However, one of the two basic assumptions, *structure assumption*, which is an essential point of graph-based SSL, is not embedded into the pairwise similarity measure. Accordingly, we propose a novel graph-based SSL method for video annotation, named Anisotropic Manifold Ranking (AniMR), based on a structure-related similarity measure. This method takes the influence of the density difference between samples into account to improve the pairwise similarity. Furthermore, we will show that AniMR can also be deduced from partial differential equation (PDE) based anisotropic diffusion. It demonstrates that the label propagation in AniMR is anisotropic, which is intrinsically different from the isotropic label propagation process in general graph-based SSL methods. Experiments conducted on the TRECVID data set show this approach outperforms ordinary graph-based SSL methods and is effective for video semantic annotation.

1. INTRODUCTION

Automatic annotation (also called as high-level feature extraction in TRECVID [1]) of video and video segments is essential for enabling semantic-level video search. As manually annotating large video archive is labor-intensive and time-consuming, efficient automatic annotation methods are desired. To this end, generally statistical models are built from manually pre-labeled samples, and then the labels are automatically assigned to the unlabeled samples using these models. However, this process has a major obstacle: frequently the labeled data is limited so that the distribution of the labeled data typically can not well represent the distribution of the entire data set (including labeled and unlabeled), which usually leads to inaccurate annotation results.

Semi-supervised learning (SSL), which attempts to learn from both labeled and unlabeled data, is a promising approach to deal with the above issue. As a major family of SSL, graph-based methods have attracted more and more researchers' attention recently. Many works on this topic are reported in the literature of machine learning community [2][12][13][14] and some of them have been applied to image or video semantic annotation. In [4], manifold-ranking is applied to propagate label information of image samples
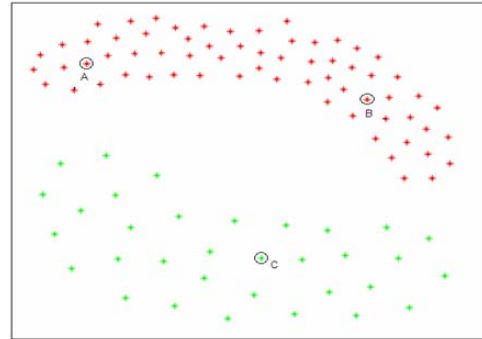
Fig. 1. Example for the influence of *structure assumption*

from positively-labeled ones to unlabeled ones. Wang *et al.*[9] propose a method based on random walk with restarts to refine the image annotation results. In [11], a manifold ranking method based on feature selection is proposed for video concept detection. Tang *et al.*[10] embed the temporal consistency of video data into the graph based SSL and propose a temporally consistent Gaussian random field method for video annotation.

Graph-based SSL actually relies on two basic assumptions [12]: *neighborhood assumption*: nearby points are likely to have the same label; *structure assumption*: points on the same "structure" (typically referred to as a cluster or a manifold) are likely to have the same label. Pairwise similarity measure is essential for graph-based methods, as it is the basis of label propagation. The first assumption is easily to be enforced into the pairwise similarity as

$$w_{ij} = \begin{cases} exp(-\frac{||x_i-x_j||^2}{2\sigma^2}) & i \neq j \\ 0 & i = j \end{cases}. \tag{1}$$

However, the second assumption is not taken into account in this typical similarity definition. Instead it is generally combined into most methods by an iterative label propagation process, but the propagation is still "isotropic" though weighted by distance. It means that the direct contribution from one sample to another, which is proportional to the pairwise similarity, neglects the influence of the structure difference. We believe that embedding the *structure assumption* into the pairwise similarity will further improve the performance of normal graph based SSL methods. Fig.1 shows an exemplary case, where each point represents a sample in the feature space. The distance between A and C is equal to the distance between A and B, that is, w_{AC} equals to w_{AB} according to the normal similarity measure defined in (1). However, it is more reasonable if the similarity w_{AB} is larger than w_{AC}, as A and B are on the same structure while A and

C are on different structures. According to the *structure assumption*, it is necessary to prevent the propagation between different clusters. That is to say, it is more appropriate to make the direct propagating strength between samples in one cluster stronger than the counterpart in different clusters.

In this paper, we embed the *structure assumption* into the similarity measure and propose a novel graph-based SSL method named Anisotropic Manifold Ranking (AniMR) for video annotation. Furthermore, we will show that actually AniMR can be deduced from a partial differential equation (PDE) based anisotropic diffusion framework [7][8]. From the view of PDE based diffusion, we can see that the label propagation in AniMR is anisotropic, which is intrinsically different from the isotropic label propagation process in general graph-based methods. That is why we call our method Anisotropic Manifold Ranking.

2. ANISOTROPIC MANIFOLD RANKING

Let $X = \{x_1, x_2, ..., x_l, x_{l+1}, ..., x_n\}$ be a set with n samples (i.e., video shots for our application) in R^m (m-D feature space). The first l samples are labeled as $\mathbf{f}_L = [f_1, f_2, ..., f_l]^T$ with $f_i \in \{1, 0\}$ ($1 \leqslant i \leqslant l$) and the remaining samples $x_u (l + 1 \leqslant u \leqslant n)$ are unlabeled. Consider a connected undirected graph (V, E) with the vertex set V corresponding to the n data points. $V = L \bigcup U$, where the vertex set $L = \{1, ..., l\}$ contains labeled points and the vertices in set $U = \{l+1, ..., l+u\}$ are unlabeled. The edges E are weighted by the $n \times n$ pairwise similarity matrix.

As aforementioned, the similarity measure will be more accurate if embedding the *structure assumption* into it. It is intuitive that generally the density variation within a cluster is smaller than the density variation between different clusters. We assume that the similarity between two samples, not only decreases with respect to the increment of their distance in the feature space, but also decreases with the increment of their density difference. This pairwise similarity is more consistent with the *structure assumption*.

Define a density similarity matrix G with entries:

$$g_{ij} = exp(-\frac{(p_i - p_j)^2}{2\sigma_p^2}), \tag{2}$$

where p_i is the probability density of sample x_i. And the new structure-related similarity matrix is defined as

$$\tilde{W} = W \bullet G, \tag{3}$$

where $\bullet$ represents the *Hadamard product* [5] and W is the normal similarity matrix with entries w_{ij}, defined in (1). Therefore, the entry of i-th row and j-th column in $\tilde{W}$ is

$$\tilde{w}_{ij} = w_{ij} \cdot g_{ij} = exp(-\frac{||x_i - x_j||^2}{2\sigma^2}) \cdot exp(-\frac{(p_i - p_j)^2}{2\sigma_p^2}) \quad (4)$$

while $i \neq j$ and $\tilde{w}_{ii} = 0$. The first term in the right side of (4) shows that the similarity between two samples decreases with respect to the increment of their distance in the feature space; and the second one indicates that the similarity decreases with the increment of the density difference. That is to say, this similarity definition not only considers the *neighborhood assumption*, but also takes the *structure assumption* into account.

Our objective is to obtain the real-valued labels for the unlabeled samples. Motivated by the two basic assumptions, we obtain the prediction function by minimizing the energy function

$$E(f) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} g_{ij} (f_i - f_j)^2, \tag{5}$$

subjects to the invariant constraint for the labels of the labeled data, that is $\mathbf{f}_L^* = \mathbf{f}_L$. Similar to [14], we can see the minimum energy function $f^* = argmin_{\mathbf{f}_L^* = \mathbf{f}_L} \tilde{E}(f)$ is harmonic. That is to say, it satisfies $\tilde{\Delta} f = 0$ on the unlabeled samples. Here $\tilde{\Delta}$ is the *combinatorial Laplacian* [3], represented with matrix form as $\tilde{\Delta} = \tilde{D} - W \bullet G$, where $\tilde{D} = diag(\tilde{d}_i)$ is the diagonal matrix with entries $\tilde{d}_i = \sum_{j=1}^{n} w_{ij} g_{ij}$.

According to the harmonic property, the value of f wrt any sample is the weighted average of the values wrt its neighboring samples. That is,

$$f_i = \frac{1}{\sum_{j=1}^{n} w_{ij} g_{ij}} \sum_{j=1}^{n} w_{ij} g_{ij} f_j = \frac{1}{\tilde{d}_i} \sum_{j=1}^{n} \tilde{w}_{ij} f_j. \tag{6}$$

Then the direct propagating strength from sample x_i to x_j is:

$$\tilde{s}_{ij} = \frac{\tilde{w}_{ij}}{\tilde{d}_i} = \frac{w_{ij} g_{ij}}{\sum_{j=1}^{n} w_{ij} g_{ij}} \tag{7}$$

It is easy to see that $\tilde{s}_{ij} \geqslant 0$, $\tilde{s}_{ii} = 0$ and $\sum_{j=1}^{n} \tilde{s}_{ij} = 1$.

Denote the propagation matrix as $\tilde{S} = \tilde{D}^{-1}(W \bullet G)$ and the predicted label vector of the entire data set with $\mathbf{f}$. Combine (6) and (7), and represent it in matrix manner, we have:

$$\mathbf{f} = \tilde{S}\mathbf{f}. \tag{8}$$

Split the matrix $\tilde{S}$ after the l-th row and l-th column:

$$\tilde{S} = \begin{bmatrix} \tilde{S}_{LL} & \tilde{S}_{LU} \\ \tilde{S}_{UL} & \tilde{S}_{UU} \end{bmatrix}.$$

Also $\mathbf{f}$ can be split into 2 blocks after the l-th row:

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_L \\ \mathbf{f}_U \end{bmatrix}.$$

Then (8) will be transformed to:

$$\begin{cases} \mathbf{f}_L = \tilde{S}_{LL}\mathbf{f}_L + \tilde{S}_{LU}\mathbf{f}_U \\ \mathbf{f}_U = \tilde{S}_{UL}\mathbf{f}_L + \tilde{S}_{UU}\mathbf{f}_U \end{cases}$$

Enforce the constraint $\mathbf{f}_L^* \equiv \mathbf{f}_L$ and solve the second linear equation, consequently we obtain the optimal solution for $\mathbf{f}_U$:

$$\mathbf{f}_U^* = (I - \tilde{S}_{UU})^{-1} \tilde{S}_{UL} \mathbf{f}_L. \tag{9}$$

In next section, we will show that this result also can be deduced in an iterative manner from a novel viewpoint, that is, PDE based diffusion. We will also show that the label propagation in this method is anisotropic while the propagation in the normal graph-based SSL methods is isotropic.

3. ANALYSIS FROM THE VIEW OF DIFFUSION

The discrete version of the heat diffusion equation [8], $\frac{\partial f}{\partial t} = \Delta f$, where Δ is the *combinatorial Laplacian*, can be written as:

$$f_i^{(t+1)} - f_i^{(t)} = \frac{1}{d_i} \sum_{j \neq i} w_{ij}(f_j^{(t)} - f_i^{(t)}). \tag{10}$$

Using the matrix representation, we have:

$$\mathbf{f}^{(t+1)} - \mathbf{f}^{(t)} = (D^{-1}W - I)\mathbf{f}^{(t)},$$

493

and therefore

$$\mathbf{f}^{(t+1)} = D^{-1}W\mathbf{f}^{(t)} = S\mathbf{f}^{(t)},$$

where $W = [w_{ij}]_{n \times n}$ is the similarity matrix with $w_{ii} = 0$, S is the normalization of W with the entries $s_{ij} = \frac{w_{ij}}{d_i}$ and $D = diag(d_i)$ is the diagonal matrix with entries $d_i = \sum_{j=1}^{n} w_{ij}$. Enforce the constraint $\mathbf{f}_L^{(t)} \equiv \mathbf{f}_L$ into the above equation, we get:

$$
\begin{aligned}
\mathbf{f}_U^{(t)} &= S_{UL}\mathbf{f}_L^{(t-1)} + S_{UU}\mathbf{f}_U^{(t-1)} \\
&= S_{UL}\mathbf{f}_L^{(t-1)} + S_{UU}(S_{UL}\mathbf{f}_L^{(t-2)} + S_{UU}\mathbf{f}_U^{(t-2)}) \\
&= ... \\
&= (\sum_{i=0}^{t-1} S_{UU}^i)S_{UL}\mathbf{f}_L + S_{UU}^t\mathbf{f}_U^{(0)}
\end{aligned}
\tag{11}
$$

It is obvious that $s_{ij} \geqslant 0$, $s_{ii} = 0$ and $\sum_{j=1}^{n} s_{ij} = 1$, S and S_{UU} are both non-negative matrices. When we connect each vertex to all other vertices (the case of a sparse representation of the matrix will be discussed in Section 4), every vertex $j \in L$ will satisfy $s_{ij} > 0$ when $j \neq i$, which results in that $\sum_{j \in L} s_{ij} > 0$. Therefore we have

$$\sum_{j \in U} s_{ij} = \sum_{j=1}^{n} s_{ij} - \sum_{j \in L} s_{ij} \leqslant 1 - \sum_{j \in L} s_{ij} < 1.$$

According to the spectral diameter bound for the non-negative matrix in the matrix theory [5], that is,

$$min_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} a_{ij} \leqslant \rho(A) \leqslant max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} a_{ij}, \tag{12}$$

we have:

$$\rho(S_{UU}) \leqslant max_{l+1 \leqslant i \leqslant n} \sum_{j \in U} s_{ij} < 1. \tag{13}$$

Therefore we get $lim_{t \to \infty} \sum_{i=0}^{t-1} S_{UU}^i = lim_{t \to \infty}(I - S_{UU})^{-1}(I - S_{UU}^t) = (I - S_{UU})^{-1}$ and $lim_{t \to \infty} S_{UU}^t = \mathbf{0}$, where $\mathbf{0}$ is a $(n - l) \times (n - l)$ matrix with each entry equals to 0. Through iterations until convergence, the optimal result will be obtained:

$$\mathbf{f}_U^* = lim_{t \to \infty} \mathbf{f}_U^{(t)} = (I - S_{UU})^{-1} S_{UL}\mathbf{f}_L \tag{14}$$

This result is the same as the result of a normal graph based SSL method: *Gaussian Random Field* (GRF) [14]. Since the heat diffusion is isotropic, the label propagation procedure in this method is distance weighted isotropic, which is not accordant with the *structure assumption* of graph-based SSL.

As aforementioned, *structure assumption* encourages propagating label information within a region with uniform or close density in preference to propagating across the density boundaries. Accordingly, we consider an anisotropic diffusion equation:

$$\frac{\partial f}{\partial t} = div(g(|\nabla p|)\nabla f) \tag{15}$$

This is a higher dimensional generalization of the anisotropic diffusion equation in [7], where $g(x)$ is a nonnegative decreasing function and p is the density distribution. Discretize the equation like the Perona-Malik Discrete Formulation [7][8], we have

$$f_i^{(t+1)} - f_i^{(t)} = \frac{\gamma_i}{\sum_{j=1}^{n} w_{ij}} \sum_{j \neq i} w_{ij} g_{ij}(f_j^{(t)} - f_i^{(t)})$$

where w_{ij} is the distance weighted similarity and g_{ij} is the density similarity.

Set $\gamma_i = \sum_{j=1}^{n} w_{ij} / \sum_{j=1}^{n} w_{ij} g_{ij}$ and notice that $w_{ii} = 0$, we have:

$$f_i^{(t+1)} - f_i^{(t)} = \frac{1}{\sum_{j=1}^{n} w_{ij} g_{ij}} \sum_{j=1}^{n} w_{ij} g_{ij}(f_j^{(t)} - f_i^{(t)}) \tag{16}$$

Represent (16) using matrix form, we obtain:

$$
\begin{aligned}
\mathbf{f}^{(t+1)} &= \mathbf{f}^{(t)} + (\tilde{D}^{-1}(W \bullet G) - I)\mathbf{f}^{(t)} \\
&= \tilde{D}^{-1}(W \bullet G)\mathbf{f}^{(t)} = \tilde{S}\mathbf{f}^{(t)}
\end{aligned}
\tag{17}
$$

where $\tilde{S}$ represents the matrix whose entry in *i-th* row and *j-th* column is $\frac{w_{ij} g_{ij}}{\sum_{j=1}^{n} w_{ij} g_{ij}}$ and $\tilde{s}_{ii} = 0$. This matrix takes both the pairwise distance and density difference into consideration. Then the predicted label vector of unlabeled data will be obtained:

$$\mathbf{f}_U^{(t)} = \tilde{S}_{UL}\mathbf{f}_L^{(t-1)} + \tilde{S}_{UU}\mathbf{f}_U^{(t-1)}. \tag{18}$$

It is worthy to notice that, although (18) have the similar form with (11), they have different intrinsic meanings: the label propagation in (11) is distance weighted isotropic while the propagation in (18) is density-sensitive anisotropic.

Enforce the constraint $\mathbf{f}_L^{(t)} \equiv \mathbf{f}_L$, similar to the process above we can obtain the result:

$$\mathbf{f}_U^* = lim_{t \to \infty} \mathbf{f}_U^{(t)} = (I - \tilde{S}_{UU})^{-1}\tilde{S}_{UL}\mathbf{f}_L \tag{19}$$

Consequently, we deduced the same result of AniMR from the viewpoint of anisotropic diffusion. The deep researches in PDE based diffusion [8] will give us new insights for graph-based SSL.

4. IMPLEMENTATION ISSUE

In the AniMR algorithm, we have to calculate the inversion or the multiplication of the large-scale matrices in (9) or in the iterations of (18), which are difficult to be implemented subject to the limitation of both the computing ability and the memory quantity. For example, the video data set typically is very large (e.g., TRECVID05 dataset has about 126,000 sub-shots); it is difficult to storage the similarity matrix and compute its inversion. To deal with this issue, we simplify the graph by only connecting neighboring points, thus matrices $\tilde{W}$ and $\tilde{S}$ are sparse, which are calculated off-line. In this way, the quantity of memory and the processing time requested are greatly reduced.

There are two methods to find appropriate set of neighboring points for calculating the sparse representation of the matrices $\tilde{W}$ and $\tilde{S}$ [6]: (a), k-NN: find the k nearest neighbors for each point; and (b), ϵ-NN: find nearest neighbors in the super-sphere centered at current point with radius of ϵ.

Another issue is to guarantee the convergence of the iterations. As aforementioned, for each sample x_i, $\sum_{j \in U} \tilde{s}_{ij} < 1$ a necessary condition to ensure the convergence of the iteration process (18). But when we use k-NN or ϵ-NN to choose the connecting neighbors, this requirement may not be satisfied since the chosen neighbors may all belong to the unlabeled set U, then $\sum_{j \in U} \tilde{s}_{ij}$ will equal to 1. That is to say, we cannot ensure the convergence of the iterative process in this case.

To tackle this difficulty, we introduce a degradation factor λ ($\lambda = 1 - \delta$, where δ is a small arithmetic number) into (17) as:

$$\mathbf{f}^{(t+1)} = \lambda \tilde{S} \mathbf{f}^{(t)}.$$

Subject to the invariant constraint $\mathbf{f}_L^{(t)} \equiv \mathbf{f}_L$, we obtain:

$$
\begin{aligned}
\mathbf{f}_U^{(t)} &= \lambda \tilde{S}_{UL} \mathbf{f}_L^{(t-1)} + \lambda \tilde{S}_{UU} \mathbf{f}_U^{(t-1)} \qquad (20) \\
&= \sum_{i=0}^{t-1} (\lambda \tilde{S}_{UU})^i (\lambda \tilde{S}_{UL}) \mathbf{f}_L + (\lambda \tilde{S}_{UU})^t \mathbf{f}_U
\end{aligned}
$$

It have been shown that $\sum_{j \in U} \tilde{s}_{ij} \leqslant 1$ and $0 < \lambda < 1$, which easily result in $\rho(\lambda \tilde{S}_{UU}) < 1$. This leads to $\lim_{t \to \infty} \sum_{i=0}^{t-1} (\lambda \tilde{S}_{UU})^i = (I - \lambda \tilde{S}_{UU})^{-1}$ and $\lim_{t \to \infty} (\lambda \tilde{S}_{UU})^t = \mathbf{0}$. Therefore we have:

$$\mathbf{f}_U^* = \lim_{t \to \infty} \mathbf{f}_U^{(t)} = (I - \lambda \tilde{S}_{UU})^{-1} (\lambda \tilde{S}_{UL}) \mathbf{f}_L. \qquad (21)$$

In the applications with large-scale data, we can use (20) to replace (18) to implement AniMR.

5. EXPERIMENTS

In the following experiments, we use the video corpus of TRECVID 2005, which is consisted of about 170 hours of TV news videos from 13 different programs in English, Arabic and Chinese. After automatic shot boundary detection, the development (DEV) set and the evaluation (EVAL) set contain 43907 and 45766 shots, respectively. Some shots are further segmented into sub-shots, and there are 61901 sub-shots for DEV and 64256 for EVAL set respectively.

The high-level feature extraction task is to detect the presence or absence of 10 predetermined benchmark concepts in each shot in the EVAL set. The 10 semantic concepts are *walking_running, explosion_fire, maps, flag-US, building, waterscape_waterfront, mountain, prisoner, sports* and *car* with concept IDs $1038 \sim 1047$. For each concept, systems are required to return ranked-lists of up to 2000 shots, and system performance is measured via non-interpolated mean average precision (MAP), a standard metric for document retrieval.

The low level features we used here are 225-D block-wise color moments in LAB color space, extracted over 5×5 fixed grid partitions, where each block is described by a 9-D feature.

Using the AniMR method, the 64256 sub-shots are labeled as $f(subshot_i)$, and the sub-shots in the same shot are merged using the maximum rule:

$$f(shot_m) = max_{subshot_i \in shot_m}(f(subshot_i)) \qquad (22)$$

Then the shots are ranked according to $f(shot_m)$.

The annotation evaluation results compared with SVM and two popular graph-based SSL methods (GRF [14] and *consistency method* [12]) are shown in Fig.2. The density is estimated by the distance weighted Parzen window [6]. The parameters in these methods are all tuned to be nearly optimal through 5-fold cross validations while λ is empirically set to 0.99. Comparing the results, we can see that: AniMR outperforms SVM for the concepts of *maps, flag-US, building, waterscape_waterfront, mountain, prisoner* and *sports*. It exceeds GRF for all concepts; and surpasses *consistency method* for *walking_running, maps, flag-US, building, prisoner, sports* and *car*. The MAP of AniMR is 0.252, which has an improvement of 9.6%, 10.6% and 4.8% over SVM, GRF and *consistency method* respectively. These comparisons demonstrate that AniMR improves the ordinary graph based SSL methods and is effective for video semantic annotation.
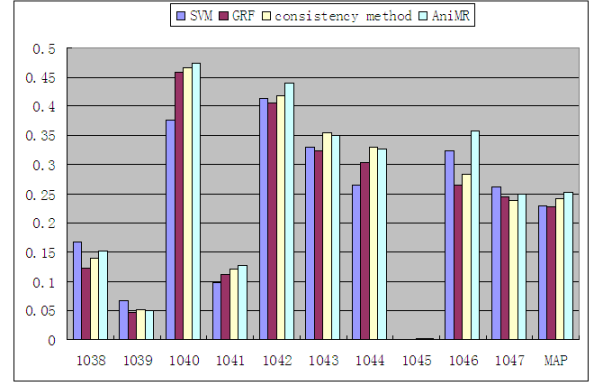

Fig. 2. The results of the 10 concepts and the MAP

6. CONCLUSION AND FUTURE WORK

In this paper, we have introduced a novel graph-based semi-supervised learning method named AniMR by embedding *structure assumption* into the pairwise similarity measure, and applied it to video semantic annotation. Furthermore, we have analyzed the proposed method from a novel viewpoint, PDE based anisotropic diffusion, which demonstrates the intrinsic difference between AniMR and normal graph-based SSL methods. Experiments conducted on the TRECVID data set demonstrate that this method outperforms the ordinary graph-based methods and effective for video annotation. Our future work is to further exploit the influence of density difference from the view of PDE based anisotropic diffusion and to find more effective solutions.

7. REFERENCES

[1] Guidelines for the TRECVID 2005 Evaluation. http://www-nlpir.nist.gov/projects/tv2005/tv2005.html.

[2] O. Chapelle, A. Zien, B. Scholkopf. *Semi-supervised Learning*, MIT Press, 2006.

[3] F.R.K. Chung. *Spectral Graph Theory*, American Mathematical Society, 1997.

[4] J. He, M. Li, et al. "Generalized Manifold-Ranking Based Image Retrieval", *IEEE Trans. Im. Proc.*, Oct. 2006.

[5] R. A. Horn, C. R. Johnson. *Matrix Analysis*, Cambridge University Press (Reprint Edition), 1999.

[6] R.O. Duda, D.G. Stork, P.E. Hart. *Pattern Classification*, JOHN WILEY, 2nd edition, 2000.

[7] P. Perona, J. Malik. "Scale-Space and Edge Detection Using Anisotropic Diffusion", *IEEE Trans. PAMI*, 12(7), 1990.

[8] G.Sapiro. *Geometric Partial Differential Equation and Image Analysis*, Cambridge University Press, 2001.

[9] C. Wang, F. Jing, et al. "Image Annotation Refinement Using Random Walk with Restarts", *Proc. ACM Multimedia*, 2006.

[10] J. Tang, X.-S. Hua, et al. "Video Annotation Based on Temporally Consistent Gaussian Random Field", *Electronics Letters*, 2007.

[11] X. Yuan, X.-S. Hua, et al. "Manifold-Ranking Based Video Concept Detection on Large Database and Feature Pool", *Proc. ACM Multimedia*, 2006.

[12] D. Zhou, O. Bousquet, et al. "Learning with Local and Global Consistency", *Proc. NIPS*, 2003.

[13] X. Zhu. *Semi-Supervised Learning with Graphs*, PhD Thesis, CMU-LTI-05-192, May 2005.

[14] X. Zhu, Z. Ghahramani, J. Lafferty. "Semi-Supervised Learning Using Gaussian Fields and Harmonic Function", *Proc. ICML*, 2003.