Semi-Supervised Manifold Ordinal Regression for Image Ranking*

Yang Liu, Yan Liu, Shenghua Zhong, Keith C.C. Chan
Department of Computing, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong, P. R. China
{csygliu,csyliu,csshzhong,cskcchan}@comp.polyu.edu.hk

ABSTRACT

In this paper, we present a novel algorithm called manifold ordinal regression (MOR) for image ranking. By modeling the manifold information in the objective function, MOR is capable of uncovering the intrinsically nonlinear structure held by the image data sets. By optimizing the ranking information of the training data sets, the proposed algorithm provides faithful rating to the new coming images. To offer more general solution for the real-word tasks, we further provide the semi-supervised manifold ordinal regression (SS-MOR). Experiments on various data sets validate the effectiveness of the proposed algorithms.

Categories and Subject Descriptors

I.5.2 [**Pattern Recognition**]: Design Methodology—*pattern analysis*

General Terms

Algorithms

Keywords

Manifold learning, ordinal regression, manifold ordinal regression, semi-supervised learning, semi-supervised manifold ordinal regression, image ranking

1. INTRODUCTION

Image ranking has attracted much attention in the last decade [11, 6, 12]. Many learning algorithms have been developed to tackle this problem, such as the regression based methods [2], the manifold learning based approaches [5, 8], and the probability based models [6].

In this paper, we present a novel algorithm called manifold ordinal regression (MOR) for image ranking, which is motivated by two observations. First, image ranking can be naturally formulated as an ordinal regression problem. Suppose that a human being is asked to rank the images according to a given query. There might be several rating scales in one's mind, such as "very relevant", "relevant", "partially relevant", or "irrelevant". Unlike the multi-class classification, ordinal regression not only recognizes whether the data points belong to the same group or not, but also provides the rank information of different data groups. Moreover, compared with the regular regression, the range of ordinal regression function is discrete and finite, which matches the rating scales in the human mind.

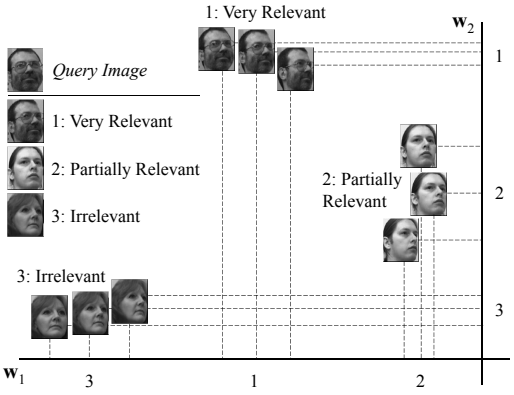

Figure 1: Schematic illustration of manifold ordinal regression on the data set with nonlinear geometry.

Second, manifold learning is an appropriate tool for image analysis. As an important category of nonlinear feature extraction technologies, manifold learning uncovers the intrinsic structure of data sets by assuming that the high-dimensional observations lie on or close to a low-dimensional manifold [14, 10, 1]. Many image data sets hold the nonlinear structures in their original high-dimensional feature spaces [14, 10, 18], which makes manifold learning being an appropriate tool for the image analysis task [5, 8, 9].

Although some effective manifold learning methods have been developed for clustering and classification [15, 9], they are not directly applicable to ordinal regression. Figure 1 provides an illustration. Given the query "man wearing glasses", we can rank the images accordingly: 1) the images of the man wearing glasses are ranked as very relevant; 2) the images of the man not wearing glasses are ranked as partially relevant; and 3) the images of the woman not wearing glasses are ranked as irrelevant. For clustering or classification, we prefer the projection that maximizes the

*Area chair: Bernard Merialdo

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

separability, so $\mathbf{w}_1$ is the optimal axis for data mapping. However, the projection on $\mathbf{w}_1$ cannot preserve the ranking information of different data blocks, which is of great importance in our task. To keep both ranking information and manifold structure, the projection on $\mathbf{w}_2$ is preferred.

According to above considerations, we proposes a novel algorithm MOR to seek the explicit function that projects the original data to the one-dimensional ranking axis. To offer more general solution for real-word tasks, we further develop the semi-supervised manifold ordinary regression (SSMOR), which learns the ordinal regression function using both labeled and unlabeled data under the proposed model.

2. MANIFOLD ORDINAL REGRESSION

Let $\{(\mathbf{x}_i, y_i)\}$ $(i = 1, ..., n)$ be the training set, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the input data point and $y_i \in \{1, ..., k\}$ denotes the corresponding rank label. In order to perform ordinal regression under the manifold learning framework, we formulate the objective function of MOR as follows:

$$\begin{aligned} &\min\ J(\mathbf{w}, \gamma) = \sum_{i,j=1}^{n} (\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}_j)^2\mathbf{A}_{ij} - C\gamma \\ &s.t.\ \ \mathbf{w}^T(\mathbf{m}_{r+1} - \mathbf{m}_r) \geq \gamma,\ r = 1, ..., k-1, \end{aligned} \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^d$ is the projection vector, $\mathbf{A}$ is the $n \times n$ adjacency matrix constructed to model the neighborhood relationship between data points, $\mathbf{m}_r$ denotes the mean vector of samples from rank r, n_r is the number of data samples in rank r, γ is the margin between the projected means of two consecutive ranks, and $C \geq 0$ is a penalty coefficient used to balance the manifold structure and order information.

2.1 Neighborhood Graph Construction

In MOR, we use the Heat kernel [1] to assign the weight on the neighborhood graph, i.e., $\mathbf{A}_{ij} = \exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2/2\sigma)$ if $j \in \mathcal{N}_i$ and $i \in \mathcal{N}_j$; and $\mathbf{A}_{ij} = 0$ otherwise. Here $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathcal{N}_i$ is the index set of the K nearest neighbors of $\mathbf{x}_i$, and $\sigma = \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{x}_{i_K})^2/n$, where $\mathbf{x}_{i_K}$ is the Kth nearest neighbor of $\mathbf{x}_i$.

To integrate the order information into the neighborhood graph, we define $d(\mathbf{x}_i, \mathbf{x}_j) = (|y_i - y_j| + 1)||\mathbf{x}_i - \mathbf{x}_j||_2$, where $|\cdot|$ denotes the absolute value operator and $||\cdot||_2$ denotes the L_2-norm operator. Compared with the L_2 distance, the distance between data points within each rank is kept unchanged while the distance between data points from different ranks is enlarged. Furthermore, the rank difference is properly reflected by the extent of enlargement.

In the neighborhood graph construction procedure, we utilize the order information from the local perspective. In the objective function (1), we aim to maximize the distance between two consecutive ranks, which considers the order information using a global manner. By incorporating the order information from both local and global perspectives, a unified manifold learning model is formulated for ranking.

2.2 Optimization Procedure

We first rewrite (1) as follows:

$$\begin{aligned} &\min\ J(\mathbf{w}, \gamma) = \mathbf{w}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{w} - C\gamma \\ &s.t.\ \ \mathbf{w}^T(\mathbf{m}_{r+1} - \mathbf{m}_r) \geq \gamma,\ r = 1, ..., k-1, \end{aligned} \tag{2}$$

where $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$ is the data matrix, $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the $n \times n$ Laplacian matrix [1], and $\mathbf{D}$ is a diagonal matrix defined as $\mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{A}_{ij}$ $(i = 1, ..., n)$. Then we obtain the Lagrangian equation of (2):

$$\begin{aligned} L(\mathbf{w}, \gamma, \alpha) = \ &\mathbf{w}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{w} - C\gamma \\ &- \sum_{r=1}^{k-1} \alpha_r\big(\mathbf{w}^T(\mathbf{m}_{r+1} - \mathbf{m}_r) - \gamma\big), \end{aligned} \tag{3}$$

where α_r are the Lagrange multipliers which satisfy $\alpha_r \geq 0$. The necessary conditions for the optimality are:

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = 0 \ \Rightarrow\ \mathbf{w} = \dfrac{1}{2}(\mathbf{X}\mathbf{L}\mathbf{X}^T)^\dagger \displaystyle\sum_{r=1}^{k-1} \alpha_r(\mathbf{m}_{r+1} - \mathbf{m}_r), \\ \dfrac{\partial L}{\partial \gamma} = 0 \ \Rightarrow\ C = \displaystyle\sum_{r=1}^{k-1} \alpha_r, \end{cases} \tag{4}$$

where $(\mathbf{X}\mathbf{L}\mathbf{X}^T)^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{X}\mathbf{L}\mathbf{X}^T$.

Based on Eq. (4), (2) could be converted to:

$$\begin{aligned} &\min \sum_{r=1}^{k-1} \alpha_r(\mathbf{m}_{r+1} - \mathbf{m}_r)^T(\mathbf{X}\mathbf{L}\mathbf{X}^T)^\dagger \sum_{s=1}^{k-1} \alpha_s(\mathbf{m}_{s+1} - \mathbf{m}_s) \\ &s.t.\ \ \sum_{r=1}^{k-1} \alpha_r = \sum_{s=1}^{k-1} \alpha_s = C,\ \alpha_r, \alpha_s \geq 0,\ r, s = 1, ..., k-1. \end{aligned} \tag{5}$$

Above is a convex quadratic programming (QP) problem with linear constraints, which can be solved by some standard optimization algorithms. Then we can obtain $\mathbf{w}$ by substituting α_r into the first equation in (4). For any data point $\mathbf{x}$, we can determine its rank by the k nearest neighbor classifier or by the decision function $f(\mathbf{x}) = \min_{r \in \{1,...,k\}} \{r : \mathbf{w}^T\mathbf{x} - b_r < 0\}$, where b_r is defined as:

$$b_r = \begin{cases} \dfrac{\mathbf{w}^T(n_{r+1}\mathbf{m}_{r+1} + n_r\mathbf{m}_r)}{n_{r+1} + n_r} & r = 1, ..., k-1, \\ \max\limits_{i \in \{1,...,n\}} \{\mathbf{w}^T\mathbf{x}_i\} & r = k. \end{cases} \tag{6}$$

2.3 Semi-Supervised Manifold Ordinal Regression

In many multimedia applications, labeled samples are often very expensive to obtain. Meanwhile, unlabeled samples are relatively easy to get and can be very helpful in discovering the global distribution of the data set [3, 19]. As shown in Figure 2, the optimal axis for data mapping is $\mathbf{w}_3$ if we consider both labeled and unlabeled images in the learning procedure, which is different from the axis $\mathbf{w}_2$ learned using only labeled images.

To offer more general solution for the real-world tasks, we develop the semi-supervised manifold ordinal regression (SSMOR), which learns the ordinal regression function using both labeled and unlabeled data under the proposed model.

Given the training set $\{\mathbf{x}_1, ..., \mathbf{x}_l, \mathbf{x}_{l+1}, ..., \mathbf{x}_n\} \subset \mathbb{R}^d$, the first l data points $\mathbf{x}_i$ $(i = 1, ..., l)$ are labeled as $y_i \in \{1, ..., k\}$ and the remaining $n - l$ data points $\mathbf{x}_u$ $(u = l + 1, ..., n)$ are unlabeled. We minimize the following cost function [17]:

$$J(\mathbf{F}) = \frac{1}{2}\sum_{i,j=1}^{n} \|\frac{\mathbf{F}_i}{\sqrt{\mathbf{D}_{ii}}} - \frac{\mathbf{F}_j}{\sqrt{\mathbf{D}_{jj}}}\|^2\mathbf{A}_{ij} + \mu\sum_{i=1}^{n} \|\mathbf{F}_i - \mathbf{Y}_i\|^2, \tag{7}$$

where $\mathbf{Y}$ denotes the $n \times k$ label matrix with $\mathbf{Y}_{ij} = 1$ if $y_i = j$, and $\mathbf{Y}_{ij} = 0$ otherwise; $\mathbf{F}$ denotes the final label matrix;

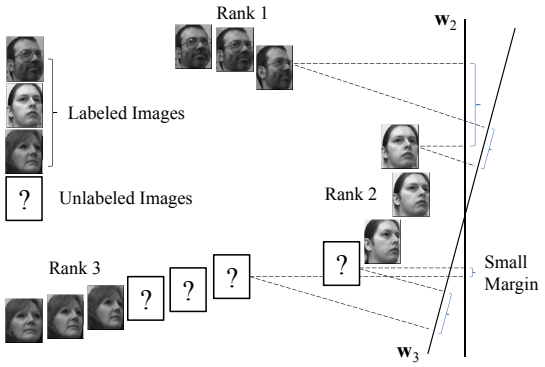

Figure 2: Schematic illustration of semi-supervised manifold ordinal regression on the data set with both labeled and unlabeled data.

$\mathbf{F}_i$ and $\mathbf{Y}_i$ denote the ith row of $\mathbf{F}$ and $\mathbf{Y}$, respectively; and $\mu > 0$ is the regularization parameter. The first term of $J(\mathbf{F})$ ensures that nearby data points have similar labels, while the second term of $J(\mathbf{F})$ requires the consistency between final labels and initial labels.

Differentiating $J(\mathbf{F})$ with respect to $\mathbf{F}$, we have

$$\frac{\partial J}{\partial \mathbf{F}} = \mathbf{F} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{F} + \mu(\mathbf{F} - \mathbf{Y}) = 0. \quad (8)$$

Then we can obtain $\mathbf{F} = \mu\big((1+\mu)\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\big)^{-1}\mathbf{Y}$. If the matrix is singular, we can use its Moore-Penrose pseudoinverse to replace the regular inverse. Finally, the label $\tilde{y}_i$ of each data point $\mathbf{x}_i$ $(i = 1, ..., n)$ can be assigned by $\tilde{y}_i = \text{argmax}_{j=1,...,k}\, \mathbf{F}_{ij}$.

According to the new labels of originally unlabeled data, we can calculate the new mean vector of each class: $\tilde{\mathbf{m}}_r = \frac{1}{\tilde{n}_r}\sum_{\tilde{y}_i=r}\mathbf{x}_i$, where $\tilde{n}_r$ is the number of data samples in class r (including the newly classified unlabeled data points). Replacing $\mathbf{m}_r$ with $\tilde{\mathbf{m}}_r$ in (1) - (5), we can obtain the optimal $\tilde{\mathbf{w}}$ and the new decision function: $\tilde{f}(\mathbf{x}) = \min_{r\in\{1,...,k\}}\{r : \tilde{\mathbf{w}}^T\mathbf{x} - \tilde{b}_r < 0\}$, where $\tilde{b}_r$ is defined as follows:

$$\tilde{b}_r = \begin{cases} \dfrac{\tilde{\mathbf{w}}^T(\tilde{n}_{r+1}\tilde{\mathbf{m}}_{r+1} + \tilde{n}_r\tilde{\mathbf{m}}_r)}{\tilde{n}_{r+1} + \tilde{n}_r} & r = 1, ..., k-1 \\ \max\limits_{i\in\{1,...,n\}}\{\tilde{\mathbf{w}}^T\mathbf{x}_i\} & r = k \end{cases} \quad (9)$$

3. EXPERIMENTS

In this section, we evaluate the performance of MOR and SS-MOR on three data sets: the UMIST face data set [4], the USPS digit data set [7], and the MSRA-MM image data set [16]. The mean absolute error (MAE, i.e., $\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$) [13] is used as the evaluation criterion. Here $\{\hat{y}_1, ..., \hat{y}_n\}$ denote the predicted ranks and $\{y_1, ..., y_n\}$ are the true targets. For all the statistical experiments, we repeat them for 20 times and report the average results. For simplicity, we set the nearest neighbor number $K = 10$ and the regularization parameter $\mu = 0.5$ in our experiments.

3.1 UMIST Face Data Set

To intuitively illustrate that MOR is capable of preserving both rank information and manifold structure, we conduct an experiment using the data samples from the UMIST face data set [4]. Consider the query "man wearing glasses".

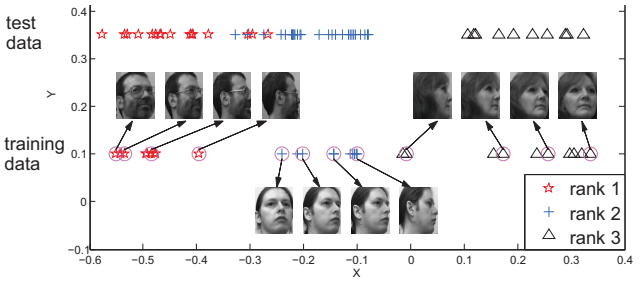

Figure 3: Projection results of training and test data on UMIST face data set using MOR.

Table 1: MAE of four algorithms on USPS data set with different training/test partition sizes.

Train/Test	MOR	KDLOR	SR	LPP
100/10900	2.73 ± 0.22	$\mathbf{2.50 \pm 0.11}$	2.85 ± 0.20	3.05 ± 0.28
200/10800	$\mathbf{2.47 \pm 0.08}$	2.48 ± 0.16	2.71 ± 0.16	2.90 ± 0.20
500/10500	$\mathbf{2.00 \pm 0.09}$	2.17 ± 0.09	2.44 ± 0.13	2.68 ± 0.12
1000/10000	$\mathbf{1.75 \pm 0.04}$	1.82 ± 0.03	2.29 ± 0.10	2.60 ± 0.11

Clearly, the images from three given classes can be ranked accordingly: 26 images of the man wearing glasses are ranked as totally match; 38 images of the man not wearing glasses are ranked as partially match; and 20 images of the woman not wearing glasses are ranked as unmatch.

Figure 3 shows the projection results of MOR. Each image is gray scale and the resolution is 56×46, so the original dimension is 2576. For each rank, 10 images are used for training and the rest are used for test. Obviously, the projected training data are arranged orderly with clear margins between consecutive ranks. Furthermore, the manifold structure within each rank, i.e., the pose variation of each person, is preserved in the projected space smoothly. For test data, although there are overlaps between rank 1 and rank 2, most of the samples are sorted correctly according to their ranks, which means that the proposed algorithm provides a faithful prediction on test data.

3.2 USPS Digit Data Set

We conduct two experiments on the United State Postal Service (USPS) data set [7]. This data set of hand written digital characters comprises 11000 normalized grayscale images of size 16×16, with 1100 images per class.

In the first experiment, our target is ranking the data according to the true digit shown in the images. We compare MOR to three representative algorithms: kernel discriminant learning for ordinal regression (KDLOR) [13], spectral regression (SR) [2], and locality preserving projections (LPP) [5]. For MOR and KDLOR, ten-fold cross validation is employed to determine the parameter C. For each class, p $(= 10, 20, 50, 100)$ images are selected for training and the rest are used for test. As shown in Table 1, MOR achieves the best results in most of the cases.

In order to demo how the performance of proposed model could be improved further when the labeled data are insufficient, in the second experiment, we test the SS-MOR by fixing the number of labeled training data and introducing the unlabeled training data. For each class, 10 images are selected as labeled training data and u $(= 10, 40, 90)$ images are selected as unlabeled training data. As shown in Table 2, the MAE of SS-MOR keeps decreasing when the number of

Table 2: MAE of SS-MOR on USPS data set with different numbers of unlabeled training data.

Labeled/Unlabeled Train	100/100	100/400	100/900
MAE of SS-MOR	2.68 ± 0.19	2.39 ± 0.16	2.21 ± 0.19

Table 3: MAE of three algorithms on MSRA-MM data set with different original feature spaces.

Features	SS-MOR	MOR	KDLOR
BWCM	$\mathbf{0.393 \pm 0.072}$	0.423 ± 0.078	0.456 ± 0.061
HSV-CH	$\mathbf{0.396 \pm 0.037}$	0.429 ± 0.019	0.440 ± 0.046
CC	$\mathbf{0.402 \pm 0.027}$	0.425 ± 0.035	0.463 ± 0.038
EDH	$\mathbf{0.379 \pm 0.026}$	0.398 ± 0.028	0.431 ± 0.063
WT	$\mathbf{0.403 \pm 0.024}$	0.421 ± 0.036	0.430 ± 0.018

unlabeled training data increases, which indicates that utilizing the unlabeled data in the training phase can improve the performance of the proposed model.

3.3 MSRA-MM Image Data Set

We conduct two experiments on a heterogeneous data set, which consists of 10 categories from the MSRA-MM image data set [16]. These categories are: baby, car, dragons, football, fruit, panda, party, school, tiger, and tree, each of which contains about 1000 images. For each image, its relevance to the corresponding query is labeled with three levels: very relevant, relevant and irrelevant.

The first experiment demonstrates the performance of the proposed MOR and SS-MOR on five different types of original feature spaces: 225-D block-wise color moment (BWCM); 64-D HSV color histogram (HSV-CH); 144-D color correlogram (CC); 75-D edge distribution histogram (EDH); and 128-D wavelet texture (WT) [16]. We compare the proposed algorithms to KDLOR, which has already shown its effectiveness in Table 1. For MOR and KDLOR, we randomly select 10 labeled images from each category for training and uses the rest for test. For SS-MOR, besides the 10 labeled images, we further use 90 randomly selected images without corresponding labels for training and uses the rest for test. Table 3 shows the MAE of SS-MOR, MOR, and KDLOR. By jointly optimizing the manifold structure and the order information, MOR outperforms the KDLOR. By incorporating the unlabeled data into the learning procedure, SS-MOR further enhances the performance.

In the second experiment, we construct the original feature space by combining aforementioned five types of features, and thus obtain a 636-dimensional vector for each image. The other settings are the same as those in the first experiment. Figure 4 shows the top ranking results generated by the SS-MOR for three categories: car, party, and tiger, respectively. The proposed algorithm obtains good results.

4. CONCLUSION

In this paper, we present a manifold ordinal regression (MOR) approach for image ranking. The first term of MOR objective function intends to preserve the manifold structure of the data set, while the second term aims to maximize the margin between the projected means of two consecutive ranks. By jointly optimizing these two terms, the projected data are orderly sorted while the intrinsic manifold structure of the data set is well preserved. To adapt the real-world applications, we present the semi-supervised manifold ordinal regression (SS-MOR). Experiments on several standard image data sets demonstrate that the proposed methods achieve good performance on the image ranking task.

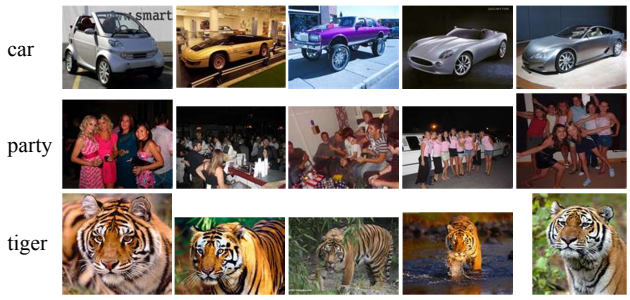

Figure 4: Top ranking results of the proposed SS-MOR for car, party, and tiger, respectively.

5. ACKNOWLEDGMENTS

This work was supported by grant PolyU 5204/09E.

6. REFERENCES

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS 14*, 2002.

[2] D. Cai, X. He, and J. Han. Spectral regression: a unified subspace learning framework for content-based image retrieval. In *Proc. 15th ACM Multimedia*, pages 403–412, 2007.

[3] O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-Supervised Learning*. MIT Press, 2006.

[4] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In *Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences 163*, pages 446–456, 1998.

[5] X. He. Incremental semi-supervised subspace learning for image retrieval. In *Proc. 12th ACM Multimedia*, pages 2–8, 2004.

[6] E. Hörster, M. Slaney, M. Ranzato, and K. Weinberger. Unsupervised image ranking. In *Proc. 1st ACM LS-MMRM Workshop*, pages 81–88, 2009.

[7] J. J. Hull. A database for handwritten text recognition research. *IEEE TPAMI*, 16(5):550–554, 1994.

[8] Y.-Y. Lin, T.-L. Liu, and H.-T. Chen. Semantic manifold learning for image retrieval. In *Proc. 13th ACM Multimedia*, pages 249–258, 2005.

[9] Y. Liu, Y. Liu, and K. C. C. Chan. Supervised manifold learning for image and video classification. In *Proc. 18th ACM Multimedia*, pages 859–862, 2010.

[10] S. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[11] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *J. Visual Commun. Image Represent.*, 10:39–62, 1999.

[12] B. Siddiquie, R. Feris, and L. Davis. Image ranking and retrieval based on multi-attribute queries. In *CVPR*, 2011.

[13] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, and W.-B. Li. Kernel discriminant learning for ordinal regression. *IEEE Trans. Knowl. and Data Eng.*, 22:906–910, 2010.

[14] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[15] H. Wang, S. Yan, T. Huang, and X. Tang. Maximum unfolded embedding: formulation, solution, and application for image clustering. In *Proc. 14th ACM Multimedia*, pages 45–48, 2006.

[16] M. Wang, L. Yang, and X.-S. Hua. Msra-mm: Bridging research and industrial societies for multimedia information retrieval. *Microsoft Technical Report*, 2009.

[17] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS 16*. 2004.

[18] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *NIPS 16*. 2004.

[19] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, pages 912–919, 2003.