Image classification with manifold learning for out-of-sample data

Article *in* Signal Processing · August 2013

DOI: 10.1016/j.sigpro.2012.05.036

4 authors, including:

Yahong Han
Tianjin University
60 PUBLICATIONS **433** CITATIONS
SEE PROFILE

Zhigang Ma
Università degli Studi di Trento
34 PUBLICATIONS **675** CITATIONS
SEE PROFILE

Zi Huang
University of Queensland
80 PUBLICATIONS **1,187** CITATIONS
SEE PROFILE

Signal Processing 93 (2013) 2169–2177

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Image classification with manifold learning for out-of-sample data

Yahong Han [a,*], Zhongwen Xu [b], Zhigang Ma [c], Zi Huang [d]

[a] School of Computer Science and Technology, Tianjin University, Tianjin, China
[b] College of Computer Science, Zhejiang University, Hangzhou, China
[c] Department of Information Engineering and Computer Science, University of Trento, Italy
[d] School of Information Technology & Electrical Engineering, The University of Queensland, Australia

ARTICLE INFO

Article history:
Received 9 December 2011
Received in revised form
21 May 2012
Accepted 28 May 2012
Available online 13 June 2012

Keywords:
Image classification
Manifold learning
Out-of-sample

ABSTRACT

The successful applications of manifold learning in computer vision and multimedia research show that the geodesic distance along the manifold is more meaningful than Euclidean distance in the linear space. Therefore, in order to get better performance of image classification, it is preferable to have classifier defined on the low-dimensional manifold. However, most of the manifold learning algorithms do not provide explicit mapping of the unseen data. In this paper, we propose a framework of image classification with manifold learning for out-of-sample data. The method of local and global regressive mapping for manifold learning simultaneously learns the low-dimensional embedding of the input data and a mapping function for out-of-sample data extrapolation. The low-dimensional manifold embedding of large-scale images can be obtained by the mapping functions. Utilizing the supervised classifier, we predict class labels for test images in the learned low-dimensional manifold. Experiments on two large-scale image datasets demonstrate that the proposed framework has better performance of image classification than the kernelized dimension reduction methods.

1. Introduction

With the development of Web 2.0 and social media applications, large-scale of image data has been uploaded on the web. For example, Facebook's servers have hosted over 100 billion of consumer photos. The explosive growth of images turns out a great challenge in image storage, indexing, and retrieval. In the preliminary stage of image retrieval, the very popular framework was to first manually annotate the images and then use the text-based retrieval technologies to perform image retrieval [1]. Due to the vast amount of labor in manual image annotation and perception subjectivity, the content-based image retrieval (CBIR) technologies [1–3] was proposed to index and retrieve images by their low-level visual features. Great deal of research works toward CBIR dedicated to finding suitable representations for images [4–6] and devising better metrics for comparisons of images [7,8].

With the extracted low-level visual features of images, many previous image retrieval techniques assume that the image space is linear and utilize the traditional Minkowski metrics to measure the similarity of images, such as the Euclidean distance. Due to the semantic gap between low-level visual features and high-level semantics, the main difficulty of such methods lies in measuring perceptual or semantical image distance by metrics in the given linear space. The latest trend in CBIR research is to find the intrinsic structure for a proper image space of reduced dimensionality. Recent studies in manifold learning show that it seems to be a reasonable assumption that the data lie in or close to a low-dimensional manifold within a high-dimensional representation space [9–13]. For example, He et al. [14] use geodesic distance to approximate the

* Corresponding author.
E-mail addresses: yahong@tju.edu.cn (Y. Han), zhongwen19901215@gmail.com (Z. Xu), kevin.z.g.ma@gmail.com (Z. Ma), huang@itee.uq.edu.au (Z. Huang).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.05.036

distances between image pairs along the manifold and apply Laplacian eigenmaps (LE) [11] to recover semantic structures hidden in the image feature space, which attained better performance of image retrieval. Therefore, the geodesic distance along the manifold is more meaningful than Euclidean distance in the linear space. In order to get better performance of image classification, it is preferable to have classifiers defined on the low-dimensional manifold [15].

When conduct manifold leaning for supervised image classification, we need first to learn the low-dimensional manifold for both the training and test images. However, different from linear dimension reduction approaches, most of the manifold learning algorithms do not provide explicit mapping of the unseen data. As a compromise, Locality Preserving Projection (LPP) [16] and Spectral Regression (SR) [17] were proposed, which introduce linear projection matrix to LE [11]. However, because a linear constraint is imposed, both algorithms are not capable to preserve the intrinsical non-linear structure of the data manifold. Furthermore, given the data dependent kernel matrix, out-of-sample data can be extrapolated for manifold learning algorithms. For example, manifold learning algorithms can be described as Kernel Principal Component Analysis (KPCA) [18] on specially constructed Gram matrices [19]. The framework proposed in [20] defined a data kernel matrix for ISOMap [9], LLE [10] and LE [11], respectively. And similar algorithm was also proposed in [21] for Maximum Variance Unfolding (MVU). One limitation of this family of algorithms is that the design of data dependent kernel matrices for various manifold learning algorithms is a nontrivial task. In [22], a nonparametric approach was proposed for out-of-sample extrapolation of LLE [10]. The low-dimensional embedding of one novel data was obtained by learning a regressive mapping from its locally linear neighbor set in training data. A limitation is that, as indicated in [22], the mapping may discontinuously change as the novel points move between different neighborhoods. Furthermore, if the novel data are off the manifold the locally linear condition is not satisfied.

Recently, a more meaningful manifold learning framework namely, Local and Global Regressive Mapping (LGRM) [23], was proposed to simultaneously learn the low-dimensional embedding of the input data and a model for unseen data extrapolation of the learned manifold. In this framework, the low-dimensional embedding of the input data was obtained by a novel Laplacian matrix. Moreover, a kernelized global regression regularization was used to extrapolate the unseen data. Based on the unified formulation [24] of the objective function of ISOMap, LLE, and LE, Yang et al. [23] discussed the connection between LGRM and other algorithms. One appealing feature of LGRM is that it not only generalizes the existing manifold learning algorithms, but also generalizes several other dimension reduction algorithms, such as KPCA [18] and KLPP [16]. Experiments on synthetic data, teapot images [25], and face expression images [10] show that LGRM preserves the manifold structure precisely and can effectively map unseen data to the learned manifold.

In this paper, we investigate the application of manifold learning for large-scale image classification with out-of-sample extrapolation. Inspired by the LGRM framework, we propose to directly learn the low-dimensional embedding of the training data, which is of moderate size. With the kernelized global regression regularization of LGRM's objective function, we can simultaneously obtain the low-dimensional mapping function for test images. Utilizing the supervised k NN classifier, we predict the class labels for the large-scale test images in the low-dimensional embedding. Experiments on two large-scale image datasets demonstrate the LGRM framework have better performance of image classification than the kernelized dimension reduction algorithms KPCA and KLPP.

The remainder of this paper is organized as follows. We first introduce related works in Section 2. The framework of local and global regressive mapping (LGRM) for manifold learning with out-of-sample extrapolation is introduced in Section 3, and we also introduce the algorithm of image classification with LGRM. The experimental analysis and conclusion are given in Sections 4 and 5, respectively.

2. Related works

2.1. Manifold learning for image/video classification

High-dimensional visual features are extracted from images. Image data reside on a low-dimensional manifold within a high-dimensional representation space. In CBIR research, recovering the intrinsic structure for a proper image space of reduced dimensionality is very important to learn a mapping function from low-level feature space to high-level semantic space [14]. Under the assumption that the data lie on a sub-manifold embedded in a high-dimensional Euclidean space, He et al. [14] proposed to conduct image analysis only on the image manifold in question rather than the total ambient space. In [26], they proposed to learn two levels of manifolds for the multimedia semantics understanding. They constructed a Laplacian media object space for media object representation and a multimedia document semantic graph to learn the semantic correlations.

Graph-based manifold learning algorithms have been well studied and applied into video classification and video concept detection. The basic idea is to propagate concept labels from labeled data to unlabeled data in the constructed graph of video shots. Traditionally, similarity of two samples is estimated by Euclidean distance between them. In order to explore the local sample and label distributions, Wang et al. [27] proposed a novel neighborhood similarity measure. The neighborhood similarity between two samples simultaneously takes into account the distribution difference of the surrounding samples and the distribution difference of surrounding labels. In image and video analysis, multiple cues should be utilized to boost the performance. OMG-SSL was proposed in [28] to integrate multiple graphs into a regularization and optimization framework. In OMG-SSL, multiple combinatorial Laplacians corresponding to multiple graphs are convexly combined. Thus, the similarities between two data samples represented by multiple

graphs are unified and taken into consideration by the combined graph Laplacians. Different from the graph Laplacians combining methodology, Zhou and Buiges [29] defined a mixture Markov model on multiple graph and gave a random walk explanation of graph Laplacians. In [30], the random walk and label propagation on multiple graphs were extended to multiple hypergraph. the method of manifold ranking on multiple hypergraph was proposed to perform video concept detection.

2.2. Manifold learning for out-of-sample data

Suppose there are n training data $\mathcal{X}=\{x_1,\ldots,x_n\}$ densely sampled from smooth manifold, where $x_i\in\mathbb{R}^d$ for $1\le i\le n$. Denote $\mathcal{Y}=\{y_1,\ldots,y_n\}$, where $y_i\in\mathbb{R}^m$ $(m<d)$ is the low-dimensional embedding of x_i. We define $Y=[y_1,\ \ldots,y_n]^T$ as the low-dimensional embedding matrix. The objective function of ISOMap, LLE, and LE can be uniformly formulated as follows [24]:

$$\min_{Y^TBY=I}\ \mathrm{tr}(Y^TLY), \tag{1}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, B is a constraint matrix, and L is the Laplacian matrix computed according to different criterions. It is easy to see that Eq. (1) generalizes the objective function of other manifold learning algorithms, such as LTSA [12]. Clearly, the Laplacian matrix plays a key role in manifold learning.

Let $x_o\in\mathbb{R}^d$ be the novel data to be extrapolated and $\mathcal{X}_o=\{x_{o1},\ldots,x_{ok}\}\subset\mathcal{X}$ be a set of data which are k nearest neighbor set of x_o in $\mathbb{R}^d$. The low-dimensional embedding y_o of x_o is given by $\sum_{i=1}^k w_i y_{oi}$, in which y_{oi} is the low-dimensional embedding of x_{oi} and w_i $(i\le i\le k)$ can be obtained by minimizing the following objective function [22]:

$$\min\ \sum_{i=1}^k \|x_o-w_iX_{oi}\|^2,\quad \text{s.t.}\ \sum_{i=1}^k w_i=1. \tag{2}$$

The formulation in Eq. (2) is a general one and can be applied to any other manifold learning algorithms for out-of-sample data extrapolation.

3. Local and global regressive mapping for image classification

3.1. Problem statement and formulation

We can construct a local clique $\mathcal{N}_i=\{x_i,x_{i,1},\ldots,x_{ik-1}\}$ for each image sample x_i, which comprises k samples, including x_i and its $k-1$ nearest neighbors [10,12,31]. We assume that the low-dimensional embedding y_j of $x_j\in\mathcal{N}_i$ can be well predicted by a local prediction function f_i, and use it to predict the low-dimensional embedding of all the data in $\mathcal{N}_i$. Since the data number in $\mathcal{N}_i$ is usually small, following Zhang and Zha [12], we perform local PCA to reduce the dimension of each datum in $\mathcal{N}_i$ as preprocessing to avoid overfitting. Because the local structure of manifold is linear [10], f_i is defined as a line regression model

$$f_i(x)=W_i^Tx+b_i,$$

where $W_i\in\mathbb{R}^{p\times m}$ is the local projection matrix and $b_i\in\mathbb{R}^m$ is bias and $p\in[m,d)$ is the dimension of each datum after local PCA being performed. To obtain a good local prediction function, we minimize the following regularized local empirical loss function [31]

$$\sum_{x_j\in\mathcal{N}_i}\|W_i^Tx_j+b_i-y_j\|^2+\gamma\|W_i\|_F^2, \tag{3}$$

where $\|\cdot\|$ denotes the Frobenius norm of a matrix.

In order to not only learn the low-dimensional embedding Y of the input data, but also learn a mapping function $f:\mathbb{R}^d\to\mathbb{R}^m$ for out-of-sample data extrapolation, we first map the data into a Hilber space $\mathcal{H}$ and assume that there is a linear transformation between $\mathcal{H}$ and $\mathbb{R}^m$ as follows:

$$y_i=\phi(W)^T\phi(x_i)+b,$$

where $\phi(W)$ is the projection matrix from $\mathcal{H}$ to $\mathbb{R}^m$ and $b\in\mathbb{R}^m$ is the bias term. We therefore propose the following objective function to simultaneously learn the low-dimensional embedding Y and the mapping function $\phi(W)\phi(\cdot)+b$:

$$\begin{aligned}\min_{\phi(W),W_i,b,b_i,Y}\ &\sum_{i=1}^n\sum_{x_j\in\mathcal{N}_i}(\|W_i^Tx_j+b_i-y_j\|^2+\gamma\|W_i\|_F^2)\\&+\mu\left(\sum_{i=1}^n\|\phi(W)^T\phi(x_i)+b-y_i\|^2+\gamma\|\phi(W)\|_F^2\right)\\ \text{s.t.}\ \ &Y^TY=I.\end{aligned} \tag{4}$$

Let $Y_i=[y_i,y_{i1},\ldots,y_{ik-1}]^T\in\mathbb{R}^{k\times m}$ be the low-dimensional embedding of $\mathcal{N}_i$. The objective function in Eq. (4) can be rewritten as

$$\begin{aligned}\min_{\phi(W),W_i,b,b_i,Y}\ &\sum_{i=1}^n(\|X_i^TW_i+1_kb_i^T-Y_i\|_F^2+\gamma\|W_i\|_F^2)\\&+\mu\left(\sum_{i=1}^n\|\phi(X)^T\phi(W)+1_nb^T-Y\|_F^2+\gamma\|\phi(W)\|_F^2\right)\\ \text{s.t.}\ \ &Y^TY=I,\end{aligned} \tag{5}$$

where $1_k\in\mathbb{R}^k$ and $1_n\in\mathbb{R}^n$ are the two vectors of all ones. Solving Eq. (5), we can simultaneously obtain the low-dimensional embedding Y of the input data and the mapping function for the out-of-sample data.

3.2. Solution and algorithm

3.2.1. Solving the low-dimensional embedding term

By employing the property that $\|M\|_F^2=\mathrm{tr}(M^TM)$ for any matrix M, the first term of Eq. (5) can be written as

$$\sum_{i=1}^n(\mathrm{tr}((X_i^TW_i+1_kb_i^T-Y_i)^T(X_i^TW_i+1_kb_i^T-Y_i))+\gamma\ \mathrm{tr}(W_i^TW_i)). \tag{6}$$

Setting derivative of Eq. (6) w.r.t. W_i to zero, we have

$$b_i=\frac{1}{k}(Y_i^T1_k-W_i^kX_i1_k). \tag{7}$$

Setting derivative of Eq. (6) w.r.t. b_i to zero, we have

$$W_i=(X_iH_kX_i^T+\gamma I)^{-1}X_iH_kY_i, \tag{8}$$

where $H_k=I-(1/k)1_k1_k^T$ is the local centering matrix. Substituted b_i and W_i by Eqs. (7) and (8), Eq. (6) can be

written as

$$\sum_{i=1}^{n} \mathrm{tr}(Y_i^T A_i Y_i), \tag{9}$$

where

$$A_i = H_k - H_k X_i^T (X_i H_k X_i^T + \gamma I)^{-1} X_i H_k. \tag{10}$$

Define a selection matrix $S \in \{0,1\}^{n\times k}$ in which $S_{ij}=1$ if x_i is the jth element in $\mathcal{N}_i$ and $S_{ij}=0$ otherwise. It is easy to see that $Y_i = S_i^T Y$ and Eq. (9) can be rewritten as

$$\mathrm{tr}\left(Y^T \left(\sum_{i=1}^{n} S_i A_i S_i^T\right) Y\right). \tag{11}$$

Let $L_l = \sum_{i=1}^{n} S_i A_i S_i^T$. Then, Eq. (11) is reformulated as [31]:

$$Y^T L_l Y. \tag{12}$$

3.2.2. Solving the mapping function term

Similarly, the second term of Eq. (5) can be written as

$$\mathrm{tr}((\phi(X)^T\phi(W) + 1_n b^T - Y)^T(\phi(X)^T\phi(W) + 1_n b^T - Y)) + \gamma\,\mathrm{tr}(\phi(W)^T\phi(W)). \tag{13}$$

Let

$$\begin{aligned} L_g &= H - H\phi(X)^T(\phi(X)H\phi(X)^T + \gamma I)^{-1}\phi(X)H \\ &= \gamma H(H\phi(X)^T\phi(X)H + \gamma I)^{-1}H \end{aligned} \tag{14}$$

then Eq. (13) can be rewritten as

$$Y^T L_g Y, \tag{15}$$

where $H = I - (1/n)1_n 1_n^T$ is the global centering matrix.

Setting derivative of Eq. (13) w.r.t. W_i to zero, we have

$$\begin{aligned} b &= \frac{1}{n}Y^T 1_n - \frac{1}{n}W^T\phi(X)1_n \\ &= \gamma\frac{1}{n}Y^T 1_n - \frac{1}{n}Y^T(H\phi(X)^T\phi(X)H + \gamma I)^{-1}H\phi(X)^T\phi(X)1_n. \end{aligned} \tag{16}$$

Setting derivative of Eq. (13) w.r.t. b_i to zero, we have

$$\begin{aligned} \phi(W) &= (\phi(X)H\phi(X)^T + \gamma I)^{-1}\phi(X)HY \\ &= \phi(X)H(H\phi(X)^T X H + \gamma I)^{-1}Y. \end{aligned} \tag{17}$$

3.2.3. The algorithm

With Eqs. (12) and (15), we have obtained the objective function of the proposed algorithm as follows:

$$\min_{Y^T Y = I} Y^T(L_l + \mu L_g)Y. \tag{18}$$

We can prove that L_l and $L_{lg} = L_l + \mu L_g$ are the Laplacian matrices. According to the unified formulation in Eq. (1), the formulation in Eq. (18) coincides with the previous manifold learning algorithm. The low-dimensional embedding Y of input data can be obtained by eigen-decomposition of $(L_l + \mu L_g)$.

For the out-of-sample data, although $\phi(X)$ cannot be explicitly computed, $\phi(X)^T\phi(X)$ can be calculated by a kernel function. We suppose that the dot production of x_i and x_j in $\mathcal{H}$ is given by the following kernel function:

$$K_{x_i,x_j} = (\phi(x_i)\cdot\phi(x_j)) = \phi(x_i)^T\phi(x_j), \tag{19}$$

where $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ can be any positive kernel satisfying Mercer's condition. In this paper, we use the Radial Basis Function (RBF) kernel, which is defined as

$$K_{x_i,x_j} = \exp(-\|x_i - x_j\|^2/\sigma^2), \tag{20}$$

where σ is a parameter. Then, L_g can be computed by

$$L_g = \gamma H(H\mathbf{K}H + \gamma I)^{-1}, \tag{21}$$

where $\mathbf{K}$ is the kernel matrix with its element $\mathbf{K}_{ij} = K_{x_i,x_j}$. Denote $\mathbf{K}_x \in \mathbb{R}^n$ as a vector with its ith element $\mathbf{K}_{xi} = \phi(x)^T\phi(x_i)$, where $x_i \in \mathcal{X}$ is the ith instance in training set. The low-dimensional embedding y of an out-of-sample data x can be computed by

$$\begin{aligned} y &= \phi(W)^T\phi(x) + b \\ &= Y^T(H\phi(X)^T\phi(X)H + \gamma I)^{-1}H\phi(X)^T\phi(x) + \frac{1}{n}Y^T 1_n \\ &\quad - \frac{1}{n}Y^T(H\phi(X)^T\phi(X)H + \gamma I)^{-1}H\phi(X)^T\phi(X)1_n \\ &= Y^T(H\mathbf{K}H + \gamma I)^{-1}H\mathbf{K}_x + \frac{1}{n}Y^T 1_n - \frac{1}{n}(H\mathbf{K}H + \gamma I)^{-1}H\mathbf{K}1_n. \end{aligned} \tag{22}$$

Algorithm 1. Image classification with Local and Global Regressive Mapping (LGRM).

Input: Matrix $X \in \mathbb{R}^{d\times n}$ of n labeled training images, where each column vector x_i ($i=1,\ldots,n$) is a d dimensional training image sample; Matrix $X_o \in \mathbb{R}^{d\times n_o}$ of n_o unlabeled test images; The number k of image samples in the local clique $\mathcal{N}_i$ for each image x_i, also of the nearest neighbors in k NN classifier; The reduced dimension m after performing local PCA for each image x_i; Parameters γ, μ, and σ.
Output: The low-dimensional embedding matrix $Y \in \mathbb{R}^{m\times n}$ and $Y_o \in \mathbb{R}^{m\times n_o}$ for training and test images; the class labels for test images.

The Training Process

construct kernel matrix $\mathbf{K}$ for training images by Eq. (19);
construct local clique $\mathcal{N}_i = \{x_i, x_{i,1}, \ldots, x_{ik-1}\}$ for each training image x_i;
construct selection matrix $S \in \{0,1\}^{n\times k}$ in which $S_{ij}=1$ if x_i is the jth element in $\mathcal{N}_i$ and $S_{ij}=0$ otherwise;
for $i=1$ to n **do**
  compute matrix A_i by Eq. (10);
end for
compute local Laplacian $L_l = \sum_{i=1}^{n} S_i A_i S_i^T$;
compute global Laplacian L_g by Eq. (21);
output Y by eigen-decomposition of $(L_l + \mu L_g)$;

The Test Process

for each image x_o in test set **do**
  for each image x_i in training set **do**
    compute $\mathbf{K}_{x_o i} = \phi(x_o)^T\phi(x_i)$;
  end for
  compute the low-dimensional embedding y_o of x_o by Eq. (22);
end for
output Y_o;
predict class labels for test images by conducting k NN classifier on Y and Y_o.

Therefore, the local and global regressive mapping (LGRM) simultaneously learns the low-dimensional embedding for input data in the training set and a mapping function for out-of-sample data. In this paper, we propose to utilize LGRM to conduct image classification. The low-dimensional embedding of the labeled training images are output directly from LGRM, whereas the low-dimensional

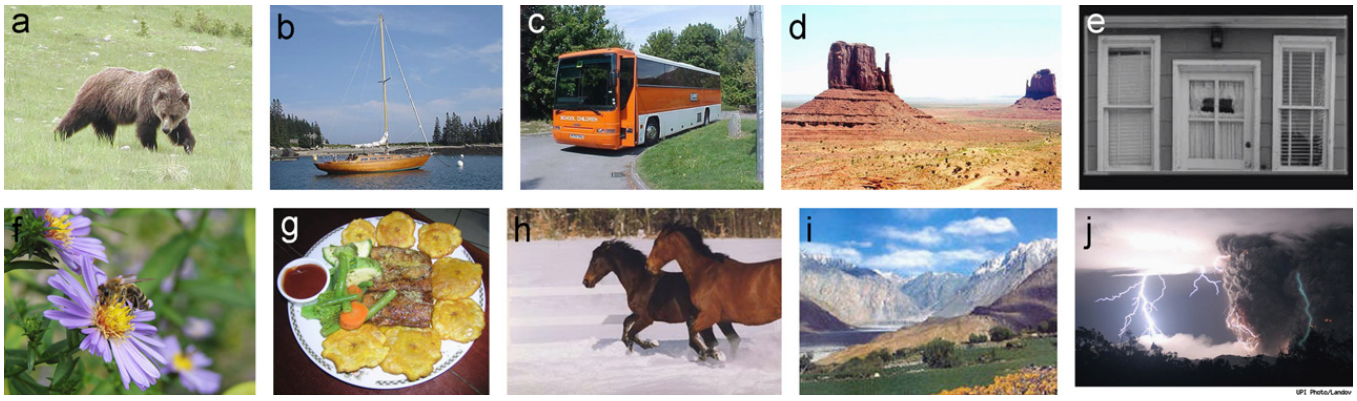

Fig. 1. Sample images and their class labels from MSRA-MM 2.0 dataset: (a) bear, (b) boat, (c) bus, (d) desert, (e) door, (f) flower, (g) food, (h) horse, (i) mountain, and (j) weather.

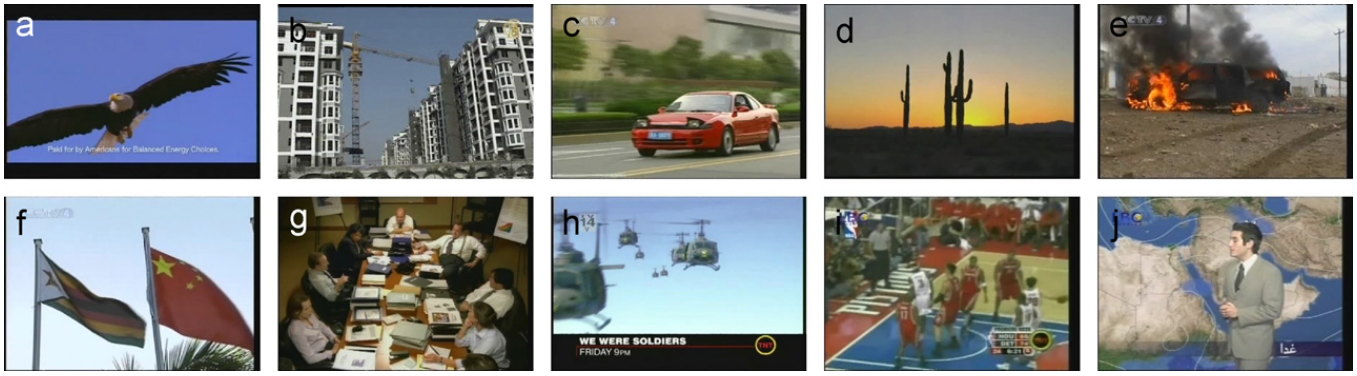

Fig. 2. Sample images and their class labels from TRECVID dataset: (a) bird, (b) building, (c) car, (d) desert, (e) explosion, (f) flag, (g) office, (h) sky, (i) sports, and (j) weather.

embedding of the unlabeled test images are obtained by the kernelized mapping function. Then we conduct supervised classification with labeled training images and test images in such low-dimensional embedding. In this paper, we use the k NN classifier for classification. We summarize the method of image classification with local and global regressive mapping in Algorithm 1.

4. Experiments

In this section, we investigate the performance of large-scale image classification with local and global regressive mapping (LGRM). Because the manifold learning algorithms, such as ISOMap [9], LLE [10] and LE [11], are incapable of out-of-sample extrapolation, we compare the proposed Algorithm 1 with two representative kernelized dimension reduction algorithms KPCA [18] and KLPP [16]. Encouraging experimental results are reported in this section.

4.1. Image datasets

Firstly, one had written digit image dataset, i.e., USPS [32], is considered as a toy example to show the performance comparison of LGRM with KPCA and KLPP.

We choose two large-scale image datsets, i.e., MSRA-MM 2.0 [33] and TRECVID,[1] in our experiments. There are tag or concept annotation for images in MSRA-MM 2.0 and TRECVID datasets, which can be taken as the ground truth for image classification. The following is a brief description of the datasets used in our experiments.

MSRA-MM 2.0: The dataset used in our experiments is a subset of the annotated images in original MSRA-MM 2.0 dataset, which includes 50,000 images related to 100 concepts. However, 7734 images within it are not associated with any labels. We have removed these images and obtained a subset of 42,266 labeled images. Four feature types, namely, Color Correlogram, Color Moment, HSV Color Histogram, and RGB Color Histogram, are combined in our experiments to represent the images. So each image sample is represented by a 689-dimension vector.

TRECVID: We use the Columbia374 baseline detectors [34] for TRECVID 2005 in our experiments. Because only images (keyframes) in the development set of TRECVID have concept annotation, the dataset used in our experiments includes 61,562 labeled images (keyframes). Three feature types used in [34], namely Edge Direction Histogram, Gabor, and Grid Color Moment, are combined in our experiments to represent the images. So each image sample is represented by a 346-dimension vector.

For USPS, we randomly sampled $n=5{,}00$ image samples to form the training set, and the rest images are test samples. For MSRA-MM 2.0 and TRECVID, we randomly sampled $n=1000$ image samples to form the training set, and the rest images are test samples. In Figs. 1 and 2, we

[1] http://www-nlpir.nist.gov/projects/trecvid/.

show some sample images and their class labels of MSRA-MM 2.0 and TRECVID datasets, respectively.

4.2. Evaluation metrics

Two typical criteria, i.e., the area under the Receiver Operating Characteristics (ROC) curve (AUC) score [35], and *F*-measure [35] are used to evaluate the performance of image classification. Since there are multiple class labels (tags or concepts) in USPS, MSRA-MM 2.0, and TRECVID datasets, to measure the global performance across multiple tags, according to Lewis [36] we use the microaveraging methods. Microaveraging values of AUC and *F*-measure are calculated by constructing a global contingency table and then calculating the measures using these sums.

4.3. Toy data example

In this experiment, we compare LGRM with KPCA and KLPP on USPS dataset. Two parameters, σ in Eq. (20) and γ, need to be set. To simplify the tuning and setting of parameters, we take γ and μ to be of the same values. We set $\sigma \in \{0.1, 1, 10, 100\}$ and $\gamma, \mu \in \{1e{-}2, 1e{-}1, 1, 10, 100\}$. For the number k of nearest neighbors of x_i used in local clique $\mathcal{N}_i$ and k NN classifier, we set $k,m = 10$, where m is the value of the reduced dimension after conducting local PCA in Algorithm 1.

Fig. 3 demonstrates the AUC and *F*-measure variance w.r.t. $\gamma(\mu)$ and σ when $k,m = 10$. We notice that the performance variance of *F*-measure is not so stable comparing with that of AUC score. Moreover, we can observe that the highest performance of the two parameters is achieved at some intermediate values.

The performance comparisons of image classification on USPS dataset are reported in Table 1. From the results, we can see that the proposed Algorithm 1 outperforms the kernelized dimension reduction methods KPCA and KLPP.

4.4. Experimental results of real images

4.4.1. Parameter sensitivity study

In this experiments, we use radial basis function (RBF) kernel (see definition in Eq. (20)) for the proposed Algorithm 1 and two comparison methods KPCA and KLPP. Therefore, the impact of parameter σ for performance of classification should be explored. Furthermore, the regularized parameters γ and μ need to be set. In order to simplify the tuning and setting of parameters, we take γ and μ to be of same values. We set $\sigma \in \{0.001, 0.01, 0.1, 1, 10, 100\}$ and $\gamma,\mu \in \{1e{-}6, 1e{-}5, 1e{-}4, 1e{-}3, 1e{-}2, 1e{-}1, 1, 10, 100, 1000\}$.

Moreover, as presented in Algorithm 1, we set the dimension of the learned low-dimensional embedding to be the same value of the reduced dimension m after conducting local PCA. Because the local PCA is used to reduce the dimension of images in the local clique $\mathcal{N}_i$, we have $m \le k$, where k is the number of nearest neighbors of x_i used in local clique $\mathcal{N}_i$ and k NN classifier. In the following, we take k, m to be of the same values. We set $k,m \in \{100, 50, 25, 10\}$.

Figs. 4 and 5 demonstrate the AUC variance w.r.t. $\gamma(\mu)$ and σ when k, m are set to be different values. In Fig. 4, it is interesting to notice that the classification performance of TRECVID is generally better when the value of σ is larger. However, the performance variance is not so stable for MSRA-MM 2.0 dataset, as shown in Fig. 5. One possible reason is that the images of MSRA-MM 2.0 are downloaded from the web by Bing image search [33], the visual content of images are usually distinguished from each other.

Table 1
The performance comparisons of image classification on USPS dataset with $k,m = 10$. The better performance are in bold face.

Parameters	Criteria	KPCA	KLPP	LGRM
$k,m = 10$	*F*-measure	0.2367	0.2265	**0.2716**
	AUC score	0.5653	0.5638	**0.5854**

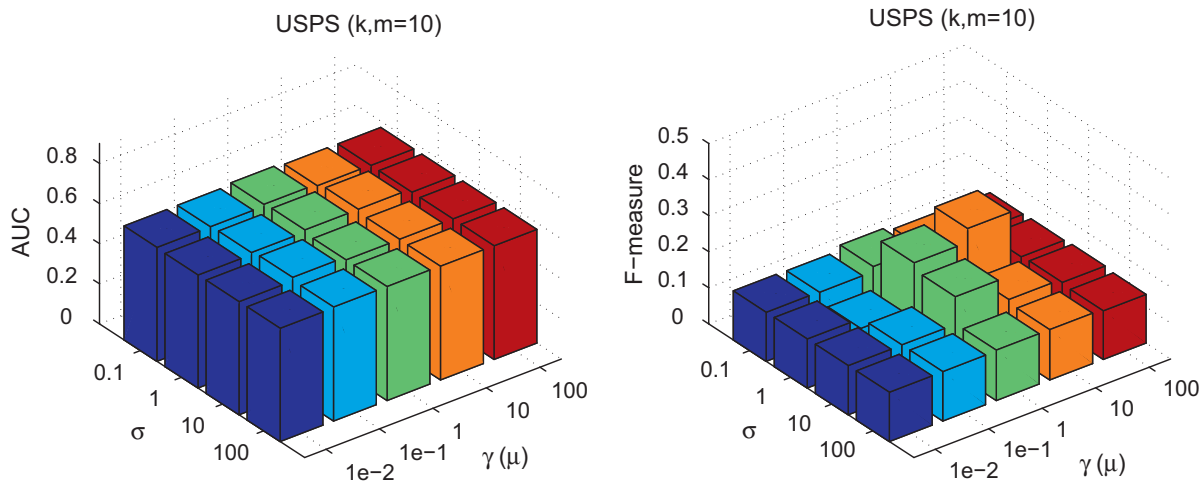

Fig. 3. Parameters sensitivity study on USPS dataset.

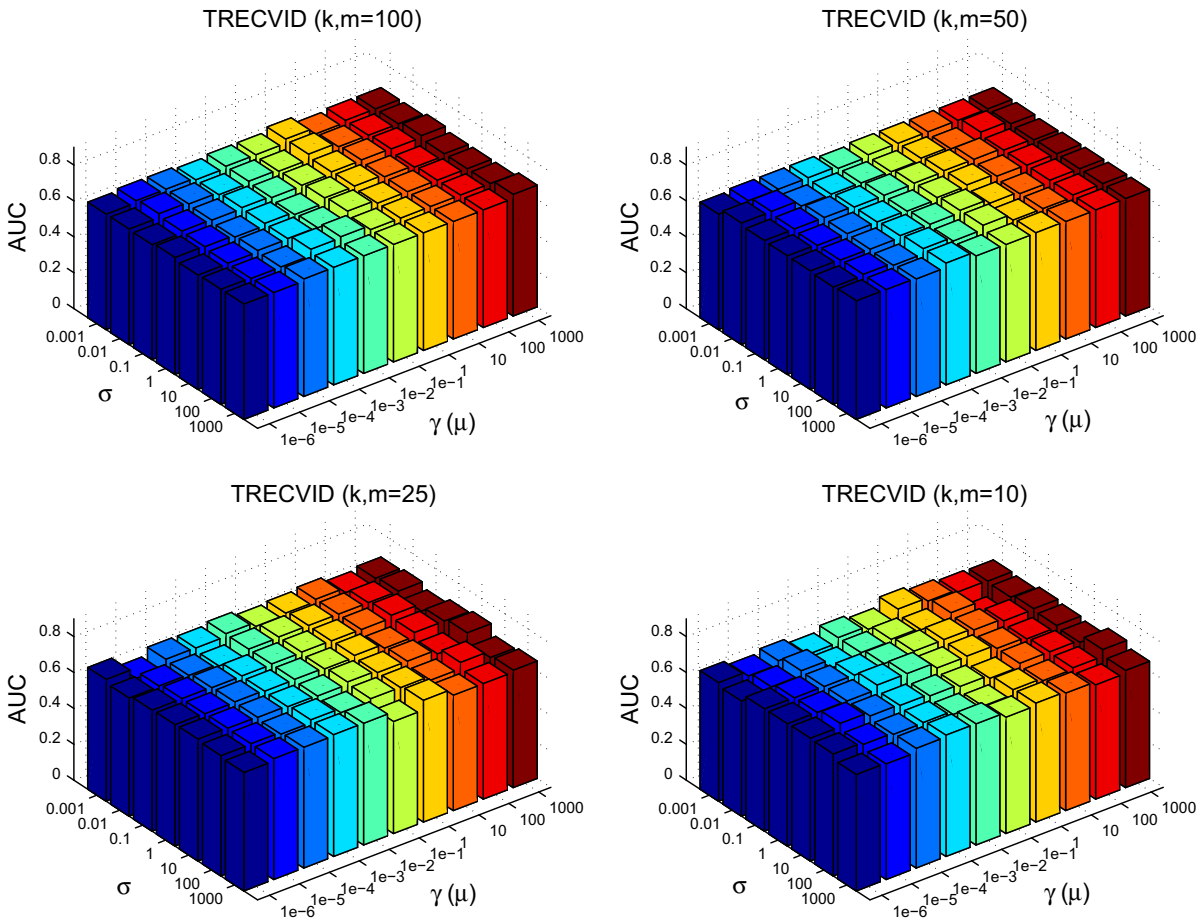

Fig. 4. Parameters sensitivity study on TRECVID dataset.

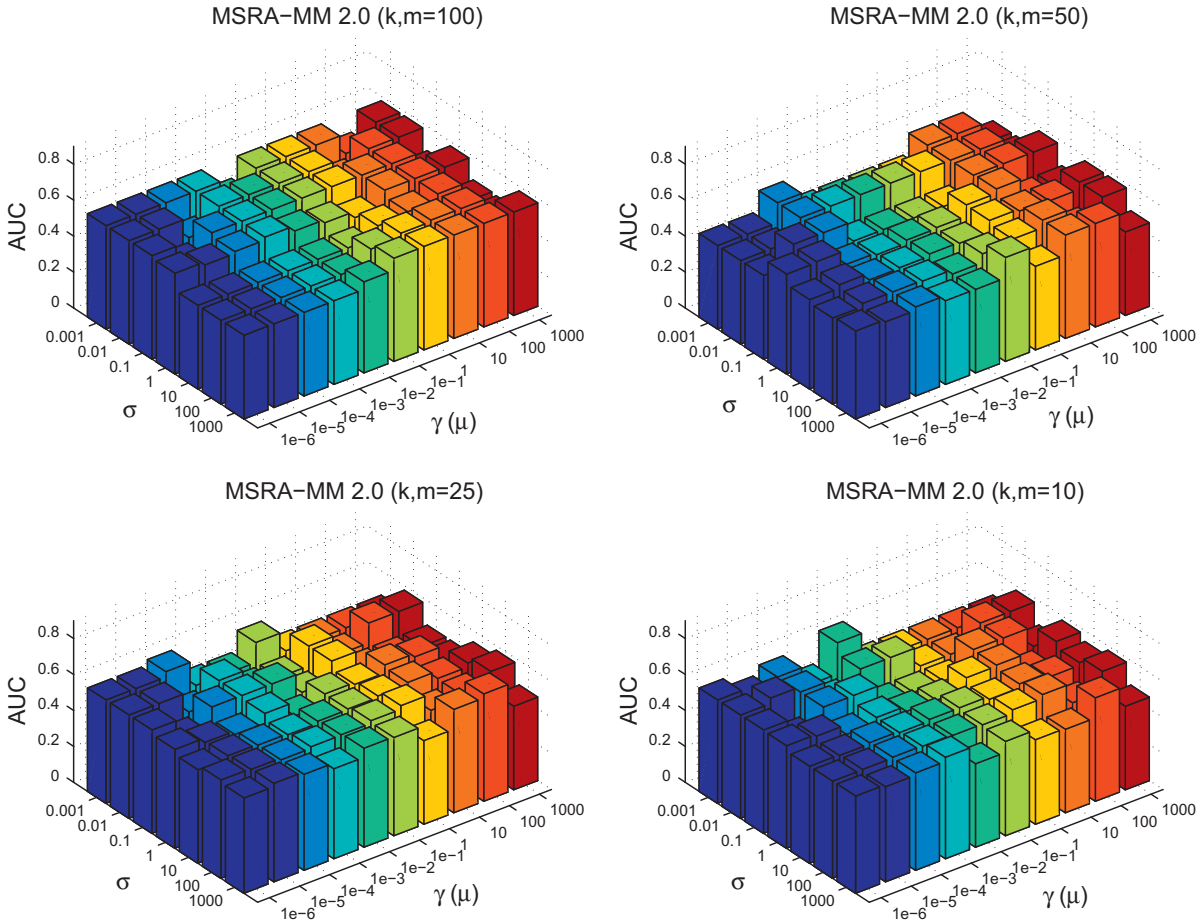

Fig. 5. Parameters sensitivity study on MSRA-MM 2.0 dataset.

4.5. Image classification results

In this section, we compare the performance of image classification of LGRM with KPCA and KLPP. According to the parameter sensitivity study in Section 4.4.1, we tune and set the parameters γ, μ, and σ of LGRM to be the values when the best performance of image classification is obtained. Similarly, the values of σ in the kernelized KPCA and KLPP are also well tuned and set with $\sigma \in \{0.001, 0.01, 0.1, 1, 10, 100\}$.

The performance comparisons of image classification on TRECVID and MSRA-MM 2.0 datasets are reported in Tables 2 and 3, respectively. Thanks to the manifold learning in LGRM, the proposed Algorithm 1 outperforms the kernelized dimension reduction methods KPCA and KLPP.

In Table 2, when k,m are set to be different values, the proposed algorithm LGRM has better image classification performance than KPCA and KLPP on TRECVID dataset. Furthermore, from the results we observe that when the number of the nearest neighbors of the local clique $\mathcal{N}_i$ and the number of the nearest neighbors of the k NN classifier are set to be the small values, we get better performance of image classification.

In Table 3, we get similar results on MSRA-MM 2.0 dataset. From the results, we observe that, when the number of the nearest neighbors of the local clique $\mathcal{N}_i$ and the k NN classifier are set to be the small values, the performance improvement of LGRM comparing with KPCA and KLPP methods is more salient. For example, when $k,m=25$ the LGRM gains about 26% improvement w.r.t. F-measure and 5% improvement w.r.t. AUC score, respectively.

Table 2
The performance comparisons of image classification on TRECVID dataset with k,m set to be different values. The better performance are in bold face.

Parameters	Criteria	KPCA	KLPP	LGRM
$k,m=100$	F-measure	0.3465	0.3235	**0.3898**
	AUC score	0.6489	0.6431	**0.6782**
$k,m=50$	F-measure	0.3686	0.3577	**0.3897**
	AUC score	0.6603	0.6545	**0.6783**
$k,m=25$	F-measure	0.3955	0.3578	**0.4133**
	AUC score	0.6754	0.6544	**0.6963**
$k,m=10$	F-measure	0.3929	0.3858	**0.4159**
	AUC score	0.6854	0.6742	**0.7054**

Table 3
The performance comparisons of image classification on MSRA-MM 2.0 dataset with k,m set to be different values. The better performance are in bold face.

Parameters	Criteria	KPCA	KLPP	LGRM
$k,m=100$	F-measure	0.2217	0.2163	**0.2319**
	AUC score	0.5651	0.5643	**0.5750**
$k,m=50$	F-measure	0.2315	0.2113	**0.2439**
	AUC score	0.5750	0.5609	**0.5822**
$k,m=25$	F-measure	0.2320	0.2317	**0.2971**
	AUC score	0.5749	0.5751	**0.6050**
$k,m=10$	F-measure	0.2764	0.2315	**0.2966**
	AUC score	0.6048	0.5751	**0.6376**

5. Conclusions

We have proposed a framework of image classification with manifold learning for out-of-sample data. The method of local and global regressive mapping (LGRM) for manifold learning simultaneously learns the low-dimensional embedding of the input data, and additionally leans a mapping function for out-of-sample data extrapolation. Therefore, we conduct large-scale image classification by first learn the low-dimensional manifold embedding for training images directly by LGRM. The low-dimensional manifold embedding for test images is obtained by the kernelized mapping function. In the reduced low-dimensional manifold, we perform supervised k NN classifier to get better performance of image classification than the kernelized dimension reduction algorithms KPCA and KLPP.

References

[1] Y. Rui, T. Huang, S. Chang, Image retrieval: current techniques, promising directions, and open issues, Journal of Visual Communication and Image Representation 10 (1) (1999) 39–62.

[2] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.

[3] R. Datta, D. Joshi, J. Li, J. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys (CSUR) 40 (2) (2008) 5.

[4] B. Manjunath, J. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11 (6) (2001) 703–715.

[5] M. Do, M. Vetterli, Wavelet-based texture retrieval using generalized gaussian density and Kullback–Leibler distance, IEEE Transactions on Image Processing 11 (2) (2002) 146–158.

[6] L. Latecki, R. Lakamper, Shape similarity measure based on correspondence of visual parts, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1185–1190.

[7] Y. Rubner, C. Tomasi, L. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision 40 (2) (2000) 99–121.

[8] J. Li, J. Wang, G. Wiederhold, IRM: integrated region matching for image retrieval, in: Proceedings of the Eighth ACM International Conference on Multimedia, ACM, 2000, pp. 147–156.

[9] J. Tenenbaum, V. Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319.

[10] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323.

[11] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems 1 (2002) 585–592.

[12] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimensionality reduction via tangent space alignment, Journal of Shanghai University (English Edition) 8 (4) (2004) 406–424.

[13] S. Xiang, F. Nie, C. Zhang, C. Zhang, Nonlinear dimensionality reduction with local spline embedding, IEEE Transactions on Knowledge and Data Engineering (2008) 1285–1298.

[14] X. He, W. Ma, H. Zhang, Learning an image manifold for retrieval, in: Proceedings of the 12th Annual ACM International Conference on Multimedia, ACM, 2004, pp. 17–23.

[15] M. Belkin, P. Niyogi, Using manifold structure for partially labeled classification, Advances in Neural Information Processing Systems (2003).

[16] X. He, P. Niyogi, Locality preserving projections, in: Proceedings of the NIPS, Advances in Neural Information Processing Systems, Vancouver, MIT Press 103.
[17] D. Cai, X. He, J. Han, Spectral regression for efficient regularized subspace learning, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2007), 2007.
[18] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299–1319.
[19] J. Ham, D. Lee, S. Mika, B. Schölkopf, A kernel view of the dimensionality reduction of manifolds, in: Proceedings of the 21st International Conference on Machine Learning, ACM, 2004, p. 47.
[20] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Le Roux, M. Ouimet, Out-of-sample extensions for LLE, ISOMAP, MDS, eigenmaps, and spectral clustering, in: Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference, vol. 16, The MIT Press, 2004, p. 177.
[21] T. Chin, D. Suter, Out-of-sample extrapolation of learned manifolds, IEEE Transactions on Pattern Analysis and Machine Intelligence (2008) 1547–1556.
[22] L. Saul, S. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, Journal of Machine Learning Research 4 (2003) 119–155.
[23] Y. Yang, F. Nie, S. Xiang, Y. Zhuang, W. Wang, Local and global regressive mapping for manifold learning with out-of-sample extrapolation, in: Proceedings of the 24th AAAI Conference on Artificial Intelligence, 2010, pp. 649–654.
[24] S. Yan, D. Xu, B. Zhang, H. Zhang, Graph embedding: A general framework for dimensionality reduction, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 2, IEEE, 2005, pp. 830–837.
[25] K. Weinberger, L. Saul, Unsupervised learning of image manifolds by semidefinite programming, International Journal of Computer Vision 70 (1) (2006) 77–90.
[26] Y. Yang, Y. Zhuang, F. Wu, Y. Pan, Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval, IEEE Transactions on Multimedia 10 (3) (2008) 437–446.
[27] M. Wang, X. Hua, J. Tang, R. Hong, Beyond distance measurement: constructing neighborhood similarity for video annotation, IEEE Transactions on Multimedia 11 (3) (2009) 465–476.
[28] M. Wang, X. Hua, R. Hong, J. Tang, G. Qi, Y. Song, Unified video annotation via multigraph learning, IEEE Transactions on Circuits and Systems for Video Technology 19 (5) (2009) 733–746.
[29] D. Zhou, C. Burges, Spectral clustering and transductive learning with multiple views, in: Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 1159–1166.
[30] Y. Han, J. Shao, F. Wu, B. Wei, Multiple hypergraph ranking for video concept detection, Journal of Zhejiang University—Science C 11 (7) (2010) 525–537.
[31] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (5) (2012) 723–742.
[32] J. Hull, A database for handwritten text recognition research, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (5) (1994) 550–554.
[33] H. Li, M. Wang, X. Hua, MSRA-MM 2.0: a large-scale web multimedia dataset, in: 2009 IEEE International Conference on Data Mining Workshops, IEEE, 2009, pp. 164–169.
[34] A. Yanagawa, S. Chang, L. Kennedy, W. Hsu, Columbia University Baseline Detectors for 374 LSCOM Semantic Visual Concepts, Columbia University ADVENT Technical Report.
[35] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (8) (2006) 861–874.
[36] D. Lewis, Evaluating text categorization, in: Proceedings of Speech and Natural Language Workshop, 1991, pp. 312–318.