Semi-Supervised Manifold Learning Based Multigraph Fusion for High-Resolution Remote Sensing Image Classification

Yasen Zhang, Xinwei Zheng, Ge Liu, Xian Sun, Hongqi Wang, and Kun Fu

***Abstract*—For high-resolution remote sensing image classification tasks, multiple features are usually required for better performances since single visual feature is valid only in describing one pattern of images. In this letter, we propose a novel Semi-Supervised Manifold learning based Multigraph Fusion framework (SSM-MF), in which multiple features are combined to learn a low-dimensional subspace. The obtained subspace can effectively characterize the semantic information of the features and thus benefits classification. Our framework employs a semi-supervised manner by exploiting labeled and unlabeled data and therefore enjoy three advancements: 1) discriminative information and geometric information in labeled data and the structural information in unlabeled data can be jointly utilized to enhance manifold learning; 2) our framework explores the complementary of multiple features and meanwhile avoids the curse of dimensionality; and 3) our semi-supervised learning mode makes use of information in abundant unlabeled data in real-world applications. Experiments on a remote sensing image data set validate the effectiveness of our proposed method.**

***Index Terms*—Manifold learning, multigraph, remote sensing image classification, semi-supervised learning.**

I. Introduction

IN THE remote sensing field, classifying images into meaningful categories is a challenging and important task, by which we can extract scene information for many environmental and socioeconomic usages. In recent years, scientists and researchers have made great effort to introduce methods from computer vision and machine learning literatures to obtain satisfying classification accuracies.

For remote sensing images, the ability of any individual feature to fully capture all aspects is limited. Therefore, multiple features are required for description from multiple views. Traditionally, feature concatenating is a common solution of combining multiple feature vectors, in which different feature vectors are concatenated into a new high-dimensional vector directly. However, such feature concatenating trick has three apparent drawbacks: 1) a concatenated vector is physically meaningless since each feature vector has a specific statistical property; 2) it ignores the diversity of multiple patterns and thus cannot efficiently explore the complementary nature of different patterns; and 3) it may result in curse of dimensionality.

Manifold learning methods assume that the instinct of high-dimensional data lies in a low-dimensional subspace. In recent years, studies show that manifold learning methods do a good job in classification tasks through learning the low-dimensional subspace. Xiao *et al.* [1] proposed a multimodal spectral clustering (MMSC) algorithm to learn a commonly shared graph Laplacian matrix by unifying different image features. However, their method is an unsupervised one. Besides, in their object function, multiple features to be equally important, which does not always agree with the fact. Zhang *et al.* [2] introduced the famous Laplacian Eigenmap (LE) into a Patch Alignment Framework (PAF) [3] to jointly unify the representation of multiple features and applied it into the hyperspectral remote sensing image classification task. The experiment results were satisfying. However, their method ignores the discriminative information contained in labeled data and structural information in unlabeled ones in the learning procedure.

In this letter, we propose a Semi-Supervised Manifold learning based Multigraph Fusion framework (SSM-MF), in which multiple features are jointly embedded into a unified low-dimensional subspace for remote sensing image classification. In SSM-MF, we perform manifold learning under PAF and then learn the joint alignment matrix. In the learning phase, discriminative information and geometric information contained in labeled data and local structural information in abundant unlabeled data is combined. Besides, by imposing optimal weights to different features, our method can explore the complementary of multiple features. The obtained low-dimensional subspace can effectively characterize the semantic concepts and has a more powerful discriminative ability.

The rest of this letter is organized as follows. Section II describes our SSM-MF framework in detail. In Section III, the SSM-MF is utilized for remote sensing image classification task and results are presented. Finally, in Section IV, we draw our conclusions.

II. Proposed SSM-MF

In this section, we present a novel Semi-Supervised Manifold learning based Multigraph Fusion framework (SSM-MF) to combine multiple features and learn a low-dimensional subspace. The learning workflow of SSM-MF is illustrated in Fig. 1. First, each remote sensing image is described using

Manuscript received November 13, 2012; revised April 4, 2013 and May 13, 2013; accepted June 4, 2013. Date of publication August 2, 2013; date of current version November 25, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 41001285 and No. 61001176.

The authors are with the Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: arsenzhang@163.com; zxw_1020@163.com; liuge86@163.com; sunxian0918@gmail.com; wiecas@sina.com; kunfu0519@sohu.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2267091

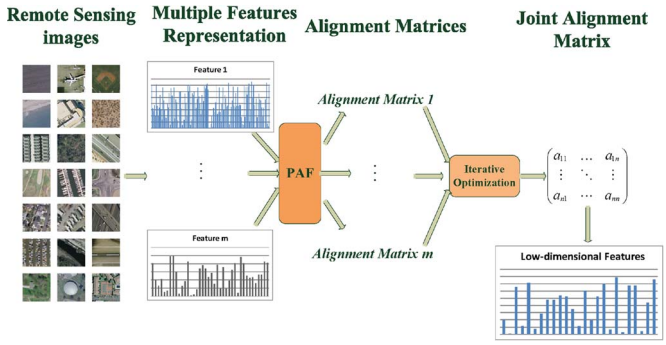

Fig. 1. Learning procedure of our framework.

multiple visual features. Second, for each feature, we obtain an alignment matrix under PAF, which was proposed in [3] as a powerful analysis tool for dimensionality reduction. PAF is composed of two stages: part optimization and whole alignment. For part optimization in SSM-MF, local patches consist of three parts: local discriminative patches, local geometric patches in labeled data and local structural patches in unlabeled data. In detail, we adopt a discriminative version of LE to calculate the local discriminative part. Locally linear embedding (LLE) [4] is utilized to preserve the local geometric relations and LE is employed by the unlabeled data to preserve the local structural information. Finally, in the whole alignment stage, weights of multiple features are learned in an iterative optimizing process and then a joint alignment matrix is obtained to calculate the final low-dimensional representations for classification usage.

Denote that each remote sensing image is described by m visual features $\{\mathbf{X}_i = [\mathbf{x}_{i,1}, \cdots, \mathbf{x}_{i,N}] \in \mathbf{R}^{V_i * N}\}_{i=1}^{m}$, in which V_i represents the dimension of the ith feature and N is the number of the images. Our proposed SSM-MF aims to find a low-dimensional representation $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N] \in \mathbf{R}^{d*N}$. In the following parts of this section, we discuss the details of our SSM-MF.

A. Part Optimization

In this section, we introduce how to optimize the three local parts.

1) Local Discriminative Patch Construction: To integrate high-level semantic information in labeled images, we firstly apply a discriminative version of LE to construct local discriminative patches using each labeled image along with its neighbors.

For $\mathbf{x}_{i,j}$, the jth sample described by the ith feature, the rests are split into two groups: samples in the same class with it and those not. From the first group, we select the k_1 nearest neighbors of $\mathbf{x}_{i,j}$ and obtain $\{\mathbf{x}_{i,j,S_1}, \cdots, \mathbf{x}_{i,j,S_{k_1}}\}$. Similarly, for the second group, we term the k_2 nearest neighbors of $\mathbf{x}_{i,j}$ as $\{\mathbf{x}_{i,j,D_1}, \cdots, \mathbf{x}_{i,j,D_{k_2}}\}$. And then, the local discriminative patch for $\mathbf{x}_{i,j}$ is described in the form of

$$\mathbf{X}_{i,j} = \{\mathbf{x}_{i,j}, \mathbf{x}_{i,j,S_1}, \ldots, \mathbf{x}_{i,j,S_{k_1}}, \mathbf{x}_{i,j,D_1}, \ldots, \mathbf{x}_{i,j,D_{k_2}}\}. \quad (1)$$

In the corresponding low-dimensional subspace, the coordinate of this local discriminative patch is represented as $\mathbf{Y}_{i,j,d} = \{\mathbf{y}_{i,j}, \mathbf{y}_{i,j,S_1}, \cdots, \mathbf{y}_{i,j,S_{k_1}}, \mathbf{y}_{i,j,D_1}, \cdots, \mathbf{y}_{i,j,D_{k_2}}\}$, which we expect to fulfill two principles: a) the sample and its k_1 nearest neighbors are as close as possible; and b) the sample and its k_2 nearest neighbors are as far as possible. Under PAF, the part optimization is formulated as

$$\arg\min_{\mathbf{y}_{i,j}} \left(\left(\sum_{a=1}^{k_1} \|\mathbf{y}_{i,j} - \mathbf{y}_{i,j,S_a}\|^2 / k_1 - \lambda \sum_{b=1}^{k_2} \|\mathbf{y}_{i,j} - \mathbf{y}_{i,j,D_b}\|^2 / k_2 \right) \right) \quad (2)$$

where λ is used to balance the influences of the aforementioned principles.

We define a coefficients vector

$$\mathbf{w}_{i,j} = [1/k_1, \ldots, 1/k_1, -\lambda/k_2, \cdots, -\lambda/k_2]. \quad (3)$$

Thus, (2) is rewritten as

$$\begin{aligned} &\arg\min_{\mathbf{y}_{i,j}} \left(\sum_{a=1}^{k_1} \|\mathbf{y}_{i,j} - \mathbf{y}_{i,j,S_a}\|^2 \mathbf{w}_{i,j,a} + \sum_{b=1}^{k_2} \|\mathbf{y}_{i,j} - \mathbf{y}_{i,j,D_b}\|^2 \mathbf{w}_{i,j,b+k_1} \right) \\ &= \arg\min_{\mathbf{y}_{i,j}} \left(\sum_{a=1}^{k_1+k_2} \|\mathbf{y}_{F_{(i,j(1))}} - \mathbf{y}_{F_{(i,j(a+1))}}\|^2 w_{i,j,a} \right) \\ &= \arg\min_{\mathbf{Y}_{i,j,d}} tr\left(\mathbf{Y}_{i,j,d} \mathbf{L}_{i,j}^{d} \mathbf{Y}_{i,j,d}^{T}\right) \end{aligned} \quad (4)$$

where the alignment matrix $\mathbf{L}_{i,j}^{d}$ is denoted as

$$\mathbf{L}_{i,j}^{d} = \begin{bmatrix} \sum_{a=1}^{k_1+k_2} \mathbf{w}_{i,j,a} & -\mathbf{w}_{i,j}^{T} \\ -\mathbf{w}_{i,j} & diag(\mathbf{w}_{i,j}) \end{bmatrix} \quad (5)$$

$F_{i,j} = \{j, j_{S_1}, \cdots, j_{S_{k_1}}, j_{D_1}, \cdots, j_{D_{k_2}}\}$ denotes the set of indices for the samples on the local patch and $diag(.)$ represents the diagonalizing operator.

2) Local Geometric Patch Construction: In the past decades, various manifold learning algorithms are proposed to recover the manifold geometry of samples in the high-dimensional space. For instance, locally linear embedding (LLE) [4] builds the embedding by preserving the geometry of pair-wise relations between samples in the high-dimensional manifold. In SSM-MF, LLE is incorporated to preserve such local geometric relations.

Using a labeled sample $\mathbf{x}_{i,j}$ and its k_3 nearest neighbor samples in the same class, we construct a geometric patch $\{\mathbf{x}_{i,j}, \mathbf{x}_{i,j,S_1}, \cdots, \mathbf{x}_{i,j,S_{k_3}}\}$. We expect that the low-dimensional representation $\mathbf{y}_{i,j}$ can be linearly reconstructed by its k_3 nearest neighbor samples, i.e.,

$$\arg\min_{\mathbf{y}_{i,j}} \left\| \mathbf{y}_{i,j} - \sum_{m=1}^{k_3} c_{i,j,m} \mathbf{y}_{i,j,S_m} \right\|^2 \quad (6)$$

in which the reconstruction coefficients $\{c_{i,j,m}\}_{m=1}^{k_3}$ can be specified by

$$\arg\min_{c_{i,j,m}} \left\| \mathbf{x}_{i,j} - \sum_{m=1}^{k_3} c_{i,j,m} \mathbf{x}_{i,j,S_m} \right\|^2, s.t. \sum_{m=1}^{k_3} c_{i,j,m} = 1. \quad (7)$$

Meanwhile, we attach the sample's k_4 nearest samples from different classes to the geometric patch and obtain

$$\begin{aligned}&\arg\min_{\mathbf{y}_{i,j}}\left\|\mathbf{y}_{i,j}-\sum_{m=1}^{k_3}c_{i,j,m}\mathbf{y}_{i,j,S_m}\right\|^2\\&=\arg\min_{\mathbf{y}_{i,j}}\left\|\mathbf{y}_{i,j}-\sum_{m=1}^{k_3+k_4}\bar{c}_{i,j,m}\mathbf{y}_{i,j,S_m}\right\|^2\\&=\arg\min_{\mathbf{Y}_{i,j,g}}tr\left(\mathbf{Y}_{i,j,g}\mathbf{L}_{i,j}^g\mathbf{Y}_{i,j,g}^T\right)\end{aligned}\tag{8}$$

in which the alignment matrix is given by

$$\mathbf{L}_{i,j}^g=\begin{bmatrix}1 & \bar{\mathbf{c}}_{i,j}^T\\ -\bar{\mathbf{c}}_{i,j} & \bar{\mathbf{c}}_{i,j}\bar{\mathbf{c}}_{i,j}^T\end{bmatrix}\tag{9}$$

$\mathbf{Y}_{i,j,g}=\{\mathbf{y}_{i,j},\mathbf{y}_{i,j,S_1},\cdots,\mathbf{y}_{i,j,S_{k_3}},\mathbf{y}_{i,j,D_1},\cdots,\mathbf{y}_{i,j,D_{k_4}}\}$ and $\bar{\mathbf{c}}_{i,j}=[c_{i,j}^T,0,\cdots,0]^T$.

3) Local Structural Patch Construction: Within the pipeline of classification, many supervised learning techniques were proposed to improve the performances. Lazebnik and Raginsky [5] incorporated discriminative information by minimizing the loss of mutual information between features and labels during the quantization step. Perronnin [6] proposed a novel practical approach based on a universal vocabulary, which describes the content of all the considered classes of images, and class vocabularies obtained through the adaptation of the universal vocabulary using class-specific data.

For classification tasks in remote sensing field, however, labels are always expensive and scarce, while unlabeled data are cheap and abundant. Recently, literatures on weakly/semi-supervised learning [7]–[9] tell that unlabeled samples are helpful to explore the local structure of distribution of the data set. The goal of LE is to find an embedding that perserves the local structure of nearby high-dimensional input patterns. In this part, we incorporate unlabeled data to construct local structural patches and thus improve the learning.

For a unlabeled data $\mathbf{x}_{i,j}^u$, we construct a local patch as $\{\mathbf{x}_{i,j},\mathbf{x}_{i,j,1}^u,\cdots,\mathbf{x}_{i,j,k_u}^u\}$, in which the k_u nearest neighbors are selected from the whole training data set. We assume that local nearby data in the high-dimensional space can stay close to each other in the low-dimensional subspace, i.e.,

$$\begin{aligned}&\arg\min_{\mathbf{y}_{i,j}}\left(\sum_{a=1}^{k_u}\|\mathbf{y}_{i,j}-\mathbf{y}_{i,j,a}\|^2\right)\\&=\arg\min_{\mathbf{Y}_{i,j,s}}tr\left(\mathbf{Y}_{i,j,s}\begin{bmatrix}-\mathbf{e}_{k_u}^T\\ \mathbf{I}_{k_u}\end{bmatrix}[-\mathbf{e}_{k_u}\mathbf{I}_{k_u}]\mathbf{Y}_{i,j,s}^T\right)\\&=\arg\min_{\mathbf{Y}_{i,j,s}}tr\left(\mathbf{Y}_{i,j,s}\mathbf{L}_{i,j}^s\mathbf{Y}_{i,j,s}^T\right)\end{aligned}\tag{10}$$

where $\mathbf{Y}_{i,j,s}=\{\mathbf{y}_{i,j},\mathbf{y}_{i,j,1},\cdots,\mathbf{y}_{i,j,k_u}\}$ and

$$\mathbf{L}_{i,j}^s=\begin{bmatrix}k_u & -\mathbf{e}_{k_u}^T\\ -\mathbf{e}_{k_u} & \mathbf{I}_{k_u}\end{bmatrix}.\tag{11}$$

B. Multigraph Whole Alignment

Describe a remote sensing data set as $\mathbf{X}=[\mathbf{x}_1,\cdots,\mathbf{x}_{N_l},\mathbf{x}_{N_l+1},\cdots,\mathbf{x}_{N_l+N_u}]$ and define a selection matrix $\mathbf{S}$ whose entry $(\mathbf{S}_i)_{pq}=\begin{cases}1 & if\ p=F_i(q),\\ 0 & otherwise\end{cases}$, where $F_i=\{i,i_1,\cdots,i_K\}$ denotes the set of indices for ith patch. The whole alignment [3] for feature i is obtained using the aforementioned three local patches

$$\arg\min_{\mathbf{Y}_i}tr\left(\mathbf{Y}_i\mathbf{L}_{T_i}\mathbf{Y}_i^T\right)\tag{12}$$

in which $\mathbf{L}_{T_i}$ is the joint alignment matrix acquired through weighted linear combination of (5), (9), and (11)

$$\begin{aligned}\mathbf{L}_{T_i}&=\sum_{j=1}^{N_l}\mathbf{S}_{i,j}\mathbf{L}_{i,j}^d\mathbf{S}_{i,j}^T+\alpha\sum_{j=1}^{N_l}\mathbf{S}_{i,j}\mathbf{L}_{i,j}^g\mathbf{S}_{i,j}^T+\beta\sum_{j=N_l+1}^{N_l+N_u}\mathbf{S}_{i,j}\mathbf{L}_{i,j}^s\mathbf{S}_{i,j}^T\\&=\mathbf{L}_i^d+\alpha\mathbf{L}_i^g+\beta\mathbf{L}_i^s.\end{aligned}\tag{13}$$

For m kinds of visual features, we fuse their graphs through a weighted alignment trick [3]

$$\arg\min_{\mathbf{Y}}\sum_{i=1}^{m}\omega_i tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T)\tag{14}$$

where ω_i represents the weight of feature i. Considering the fact that different features have different impacts on the subspace learning, we constrain that the positive weights sum to 1, and optimize them simultaneously with $\mathbf{Y}$. Moreover, to obtain the unique solution, we impose an orthogonal constraint on $\mathbf{Y}$. Thus, we have:

$$\begin{aligned}\arg\min_{\omega_i,\mathbf{Y}}\sum_{i=1}^{m}\omega_i tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T),s.t.\mathbf{Y}\mathbf{Y}^T=\mathbf{I},\omega_i>0\\ \sum_{i=1}^{m}\omega_i=1.\end{aligned}\tag{15}$$

However, a problem is encountered when optimizing (15): it is linear with respect to each ω_i, and then only one graph is kept when optimizing each ω_i. To solve this problem, we adopt an exponent trick proposed in [10], in which ω_i^r serves as a substitute for ω_i via an exponent $r>1$, i.e.,

$$\begin{aligned}\arg\min_{\omega_i,\mathbf{Y}}\sum_{i=1}^{m}\omega_i^r tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T),s.t.\mathbf{Y}\mathbf{Y}^T=\mathbf{I},\omega_i>0\\ \sum_{i=1}^{m}\omega_i=1.\end{aligned}\tag{16}$$

Since direct optimization to (16) is difficult, we adopt an iterative, two-step strategy to alternately optimize $\mathbf{Y}$ and $\{\omega_i\}_{i=1}^m$ instead.

Step 1. Fix $\mathbf{Y}$ and optimize $\{\omega_i\}_{i=1}^m$:

Using Lagrange multiplier, we introduce a Lagrange multiplier θ and rewrite the object function as

$$\arg\min_{\mathbf{Y},\omega_i,\theta}\sum_{i=1}^{m}\omega_i^r tr\left(\mathbf{Y}_i\mathbf{L}_{T_i}\mathbf{Y}_i^T\right)+\theta=\left(\sum_{i=1}^{m}\omega_i-1\right).\tag{17}$$

By setting the deviations of (17) w.r.t ω_i and θ to zero, we can obtain the weights as

$$\omega_i=\frac{\left(1/tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T)\right)^{\frac{1}{r-1}}}{\sum_{i=1}^{m}\left(1/tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T)\right)^{\frac{1}{r-1}}},\qquad i=1:m.\tag{18}$$

From (18), it can be seen that if $\mathbf{L}_{T_i}$ cannot be guaranteed to be semi positive-definite (S.P.D) or positive-definite (P.D),

$tr(\mathbf{Y}\mathbf{L}_{T_i}\mathbf{Y}^T)$ may be negative, leading w_i not to be a real value. To avoid this, we take steps to regularize $\mathbf{L}_{T_i}$ to be P.D.

Introduced by [4] and [11], both $\mathbf{L}_i^g$ and $\mathbf{L}_i^s$ are S.P.D, while $\mathbf{L}_i^d$ is neither S.P.D nor P.D. Inspired by the orthogonal constraint, we regularize $\mathbf{L}_i^d$ to $\mathbf{I}+1\eta\mathbf{L}_i^d$ to a P.D matrix by setting $\eta > max(0,-\lambda_{min})$, in which λ_{min} denotes $\mathbf{L}_i^d$'s smallest eigenvalue. It can be easily proved that this regularization has no influence on the optimization.

Step 2. Fix $\{\omega_i\}_{i=1}^m$ and optimize $\mathbf{Y}$:

Fixing $\{\omega_i\}_{i=1}^m$, the optimization problem can be expressed as: $\arg\min_{\mathbf{Y}} tr(\mathbf{Y}\mathbf{L}_T\mathbf{Y}^T), s.t. \mathbf{Y}\mathbf{Y}^T=\mathbf{I}$, where $\mathbf{L}_T = \sum_{i=1}^m \omega_i^r \mathbf{L}_{T_i}$. This optimization problem can be solved by the eigenvalue composition method. $\mathbf{Y}$ is obtained as the eigenvectors associated with the smallest non-zero eigenvalues of $\mathbf{L}_{T_i}$.

Totally, the algorithm of SSM-MF is listed in Algorithm I.

Algorithm I. SSM-MF

Input: A remote sensing data set described by m kinds of features $\mathbf{X}=\{\mathbf{X}_i=[\mathbf{x}_{i,1},\cdots,\mathbf{x}_{i,N}]\in\mathbf{R}^{V_i*N}\}_{i=1}^m$, N_l of which are labeled and the rest N_u are unlabeled.

Output: The corresponding features of $\mathbf{X}$ in the low-dimensional subspace $\mathbf{Y}=[\mathbf{y}_1,\cdots,\mathbf{y}_N]\in\mathbf{R}^{d*N}$

(1) For each feature in $\mathbf{X}$,
Calculate the alignment matrix based on discriminative patches (5), convert the alignment matrix to a positive definite matrix.
Calculate the alignment matrix based on geometric patches (9).
Calculate the alignment matrix based on structural patches (11).
Calculate the joint alignment matrix $\mathbf{L}_{T_i}$ by (13).
EndFor

(2) Solve the optimization problem of (16),
Initialize: $\omega_i=(1/m), i=1:m$.
Repeat
Perform eigenvalue composition on $\mathbf{L}_{T_i}$ and select its d eigenvectors corresponding to the d smallest non-zero eigenvalues to assemble $\mathbf{Y}$.
Update ω_i using (18).
until convergence

C. Out-of-Sample Extension for New Data

In real-world classification tasks, new data are always registered into the current data set whereas relearning a new low-dimensional subspace is quite time-consuming. Unfortunately, many manifold learning methods (e.g., LE [11], LLE [4]) suffer from the so-called out-of-sample extension (OSE) problem. Zhang *et al.* [3] assumed that a linear mapping exists between the concatenated features and the low-dimensional representations. The low-dimensional data can be obtained through the linear mapping which was precalculated in the training procedure. However, their assumption inevitably suffers from linearization errors.

Inspired by the reconstruction idea in [4], we use LLE to calculate the corresponding low-dimensional representations of new data. In detail, for a new data, we select its k_{LLE} nearest neighbors in the training set and calculate the reconstruction coefficients by LLE. Finally, the coefficients are utilized to weight the embeddings of the k_{LLE} nearest neighbors to obtain the low-dimensional representation.

III. Experiment

A. Data Set Description and Experiment Settings

To demonstrate the effectiveness of our proposed framework in remote sensing image classification, we conduct experiments on a remote sensing data set: 21 classes land-use [12], whose images were manually extracted from large images from the USGS National Map Urban Area Imagery collection for various urban areas. In this data set, each image measures 256∗256 pixels and the pixel resolution is 1 foot. Totally, this data set covers 21 classes and each class contains 100 images.

We represent each image using five kinds of visual features, including 300-dimensional SIFT-LLC [13], [14], 500-dimensional Bag-of-Features (BoF) along with Daisy [15], 512-dimensional GIST [16], 18-D Local Binary Patterns (LBP) [17] and 34-D Zernike moments. LLC is an effective coding scheme in place of vector quantization in BoF representation. GIST was proposed to model the shape of the scene and showed good performances in outdoor scene classification task. LBP is a widely-used texture feature and Zernike moments represent the global shape distributions of images.

For our proposed SSM-MF, we empirically set $k_1=k_2=k_3=5$, $k_u=10$, $\lambda=0.2$, $\alpha=0.25$ and $\beta=0.6$. The parameters d, k_{LLE} and r are selected via a five-fold cross-validation procedure. In each class, we randomly select a part as labeled samples and the rests as unlabeled ones. Meanwhile, 50% images in each class are assigned as the testing set. We randomly sample training/testing sets five times and the average classification accuracies (CA) on the testing data are reported.

The following six methods are compared with SSM-MF: Feature Concatenation (FC), Best Single Feature (BSF), Best Single Feature Embedding (BSFE), MFC [3] and two kernel fusion methods proposed in [18]: Average Kernel Fusion (AKF) and Product Kernel Fusion (PKF). In FC, all visual features are simply concatenated to a long vector. Both BSF and BSFE use single visual feature: BSF directly use the original feature while BSFE first performs embedding for each visual feature and then impose the corresponding low-dimensional representation for classification. The results of BSFE and BSF are recorded as the CA of the feature which gives the best performance. All predictions for the testing sets are carried out in SVM framework.

B. Performance Evaluation

The performances of our proposed SSM-MF, AKF, PKF, FC, BSFE, BSF, and MFC are comparatively drawn in Fig. 2 when the labeled parts are sampled as 20%, 25%, 30%, 35%, 40%, 45% and 50% of the whole respectively.

Compared with BSF, BSFE does a better job, indicating that low-dimensional features learned from our embedding framework indeed benefit the classification task. Our proposed SSM-MF is superior to the other methods. In detail, SSM-MF performs better than BSFE in that SSM-MF incorporates the complementary of multiple features by selecting optimal

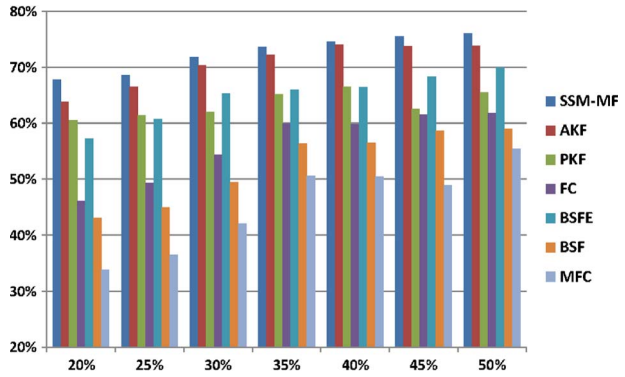

Fig. 2. Performance evaluations on classification accuracy of SSM-MF, AKF, PKF, FC, BSFE, BSF, and MFC with varied proportions as labeled samples.

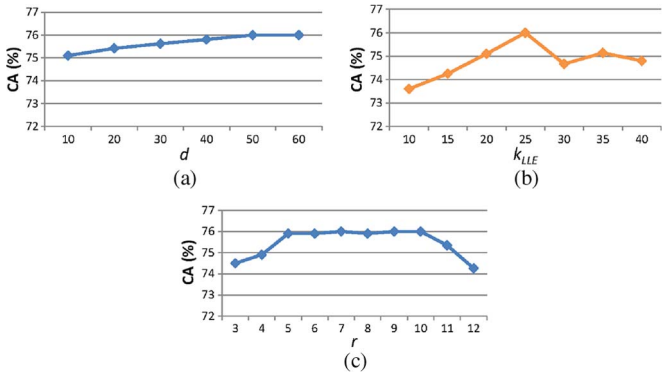

Fig. 3. Performance evaluations of parameters: (a) d; (b) k_{LLE}; and (c) r.

weights for them. Compared with SSM-MF, MFC ignores local geometric information among labeled samples and discards structural information contained in plenty of unlabeled samples. Moreover, for OSE strategy in MFC, the linearization assumption also results in considerable errors because linear relations do not always make sense between high-low dimensional features.

C. Parameters Analysis

In this section, we analyze the impact of k_{LLE}, r, and d on CA, respectively when 50% of the total samples are randomly selected as labeled samples. Also, the results reported are the average CA of five repetitions.

First, we fix k_{LLE} and r to 25 and 7, respectively and then analyze the impact of d. The curve in Fig. 3(a) tells that when d is enlarged from 10 to 40, better performances are obtained and the performances keep stable with d rising from 40 to 60.

Then, we fix d to 50 and illustrate the performances of SSM-MF with varied k_{LLE} in Fig. 3(b). From the curve, it can be seen that the optimal value of r hits 25. When $r > 25$, the CA declines in that the reconstruction error rises when using overmany samples.

Finally, d is fixed to 25 and r is evaluated. It can be seen from Fig. 3(c) that the performances stay in top level when r varies between 5 and 10. It indicates that multiple features compensate well in this range.

IV. Conclusion

In this letter, we propose a novel Semi-Supervised Manifold learning based Multigraph Fusion framework (SSM-MF) for remote sensing image classification task. In SSM-MF, multiple visual features are combined to learn a low-dimensional subspace and optimal weights are simultaneously learned to represent different impacts of these features. Furthermore, discriminative information and geometric information in labeled data and structural information in abundant unlabeled data is jointly utilized during the learning procedure. Experiments on a remote sensing data set demonstrate that our method can effectively benefit the classification tasks.

References

[1] C. Xiao, N. Feiping, H. Heng, and F. Kamangar, "Heterogeneous image feature integration via multimodal spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1977–1984.

[2] Z. Lefei, Z. Liangpei, T. Dacheng, and H. Xin, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[3] Z. Tianhao, T. Dacheng, L. Xuelong, and Y. Jie, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 22, 2000.

[5] S. Lazebnik and M. Raginsky, "Supervised learning of quantizer codebook by information loss minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1294–1309, Jul. 2009.

[6] F. Perronnin, "Universal and adapted vocabularies for generic visual categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1243–1256, Jul. 2008.

[7] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[8] G. Sharma, F. Jurie, and C. Schmid, "Discriminative spatial saliency for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 3506–3513.

[9] X. Zhu, A. B. Goldberg, R. Brachman, and T. Dietterich, *Introduction to Semi-Supervised Learning*. San Rafael, CA, USA: Morgan and Claypool Publ., 2009.

[10] W. Meng, H. Xian-Sheng, H. Richang, T. Jinhui, Q. Guo-Jun, and S. Yan, "Unified video annotation via multigraph learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 733–746, May 2009.

[11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[12] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2010, pp. 270–279.

[13] W. Jinjun, Y. Jianchao, Y. Kai, L. Fengjun, T. Huang, and G. Yihong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 3360–3367.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[15] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

[16] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, May/Jun. 2001.

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[18] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 221–228.