OBJECT MANIFOLD ALIGNMENT FOR MULTI-TEMPORAL HIGH RESOLUTION REMOTE SENSING IMAGES CLASSIFICATION

Guoming Gao [a], Meiling Zhang [a], Yanfeng Gu [a, *]

[a] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China – (gmgao01, guyf77)@gmail.com

Commission VI, WG VI/4

KEY WORDS: High spatial resolution, Segmentation, Objects, Superpixel, Manifold alignment, Multi-temporal, Classification

ABSTRACT:

Multi-temporal remote sensing images classification is very useful for monitoring the land cover changes. Traditional approaches in this field mainly face to limited labelled samples and spectral drift of image information. With spatial resolution improvement, "pepper and salt" appears and classification results will be effected when the pixelwise classification algorithms are applied to high-resolution satellite images, in which the spatial relationship among the pixels is ignored. For classifying the multi-temporal high resolution images with limited labelled samples, spectral drift and "pepper and salt" problem, an object-based manifold alignment method is proposed. Firstly, multi-temporal multispectral images are cut to superpixels by simple linear iterative clustering (SLIC) respectively. Secondly, some features obtained from superpixels are formed as vector. Thirdly, a majority voting manifold alignment method aiming at solving high resolution problem is proposed and mapping the vector data to alignment space. At last, all the data in the alignment space are classified by using KNN method. Multi-temporal images from different areas or the same area are both considered in this paper. In the experiments, 2 groups of multi-temporal HR images collected by China GF1 and GF2 satellites are used for performance evaluation. Experimental results indicate that the proposed method not only has significantly outperforms than traditional domain adaptation methods in classification accuracy, but also effectively overcome the problem of "pepper and salt".

1. INTRODUCTION

Multi-temporal remote sensing images, which are acquired by sensors mounted on board of satellites that periodically pass over the same geographical area, become an important tool for performing Earth monitoring. However, two main obstacles prevent multi-temporal technology from reaching a broader range of applications. On the one hand, there is generally a lack of labelled data at each acquisition. On the other hand, multi-temporal images obtained under different conditions show the spectral drift. Such drift generally happens due to differences in acquisition and atmospheric conditions or changes in the nature of the observed object (Tuia et al, 2016a). The obstacle of label scarcity can be solved by using available labelled samples from other temporal images. Due to spectral drift, the distributions of source image and target image are significantly different. To classify un-labelled image using labelled image efficiently and accurately, modern processing systems must be designed to be robust for solving spectral drift.

Spectral drift can be solved by shifting data distribution, which is a hot issue in machine learning referred to domain adaptation (DA) or transfer learning. In the community of multi-temporal remote sensing images classification, several approaches have been proposed to solve spectral drift. Kernel framework and manifold alignment (MA) are two kinds of typical methods. Support vector machines (SVM) was extended to the DA framework by exploiting labelled source-domain data and unlabelled target domain data in the training phase of the algorithm (Bruzzone et al, 2009). A new transfer kernel learning approach was proposed a to learn a domain-invariant kernel by directly matching the source and target distributions in the reproducing kernel Hilbert space (Long et al, 2015). Laplacian support vector machines (LapSVM) method was proposed for solving spectral drift (Kim et al, 2010). The classifier in LapSVM is adapted to the new data set via iterative application of the classifier using the clustering condition on the data manifold. Domain Transfer Multiple Kernel Learning (DTMKL) is proposed aiming at simultaneously learning a kernel function and a robust classifier by minimizing both the structural risk function and the distribution mismatch (Duan et al, 2012). MA has also been found to be useful for shifting data distributions (Yang et al, 2006; Wang et al, 2009) and overcoming spectral drift. Yang et al. proposed two MA techniques that involve aligning underlying local manifolds of temporally sequential data sets. The proposed methods extend graph-based semi-supervised learning and explore MA for the multi-temporal image classification task, while proposing a DA framework from the geometric learning viewpoint. Semi-supervised MA (SSMA) method is proposed by useing labelled samples from all domains to bring the manifolds closer while keeping their respective inherent structure unchanged using proximity graphs built with unlabelled samples (Tuia et al, 2014; Yang et al, 2016). Tuia et al. also study a generalization of SSMA through kernelization for MA in a nonlinear way. The Kernel MA (KEMA) provides a flexible and discriminative projection map, exploits only a few labelled samples in each domain, and turns to solve a generalized eigenvalue problem (Tuia et al, 2016a; Tuia et al, 2016b).

With the advent of the new generation of remote sensing missions, satellites with short revisit time and high resolution

* Corresponding author

(HR) sensors have come up and increased significantly (Tuia et al, 2016c). As a consequence, it becomes possible to perform a large variety of monitoring studies since the geographical area of interest can be covered periodically (Tuia et al, 2014). And analysts have the opportunity to use multi-temporal images for tasks such as repetitive monitoring of the territory, change detection, and large-scale processing. However, some challenges are brought to human as well as the opportunity. First, we could often find the "pepper and salt" effect in the classification results when the pixelwise classification algorithms are applied to high-resolution satellite images, in which the spatial relationship among the pixels is ignored. Second, high resolution imaging and the insufficient spectral bands make the distribution of different classes overlap seriously and reduce class separability.

For overcoming "pepper and salt" problem in high resolution images classification, superpixel segmentation is often used for keeping the consistency of adjacent space in classification. Superpixel segmentation is an important pre-processing step of many image processing algorithms, like object detection and tracking (Rasmussen, 2007), image segmentation and modelling (Mi et al, 2009), salience detection (Tong et al, 2014) and image and object classification (Liu et al, 2016), etc. A superpixel is a set of pixels which are homogeneous in perception or spectrum, a representative color or spectrum is used to represent each superpixel, and the effect of noise and distortions is thus mitigated. What's more, it can also capture image redundancy and greatly reduce the complexity of subsequent image processing tasks.

For the overlap problem existed in high spatial resolution image, tradition manifold alignment for multi-temporal classification method must improve the similarity matrix calculating step (include the in-domain and multi-domain). For classifying the multi-temporal high resolution images with limited labelled samples, spectral drift, an object-based manifold alignment method is proposed. Firstly, multi-temporal multispectral images are cut to superpixels or supervoxels by simple linear iterative clustering respectively to overcome "pepper and salt" effect (2 cases of multi-temporal high-resolution remote sensing images are considered: One is that all images cover different geographic areas, the other is that all the images cover same geographic area and are registered). Secondly, features obtained from each superpixel are formed as a vector. Thirdly, a majority voting manifold alignment is proposed to build improved graph Laplacian for mitigating overlapped phenomenon and mapping the vector data to alignment space. At last, all the data in the alignment space are classified by using KNN method.

The rest parts of this paper are organized as follows. Section 2 describes the proposed method. Section 3 presents the two groups of multi-temporal HR images used in the experiments and summarizes experimental results in details. Section 4 draws conclusions for this paper.

2. PROPOSED METHOD

Two cases of multi-temporal high-resolution remote sensing images are considered in this paper: One is that all images cover different geographic areas, another is that all the images cover the same geographic area and are registered. The first case mainly considers the spectral drift problem between different images and has a wide range of application in multi-temporal, multi-source and multi-model data fusion and classification. The second case mainly considers the continuity of segmentation results of unchanged areas between multi-temporal images, excepting considering the spectral drift. This case has a deeply using in high resolution images change detection. For the first case, a new object based multi-temporal high resolution classification methodology is proposed and consists of four steps: respective multi-temporal images segmentation (cut to superpixel, treat as object), object feature extraction, object alignment with majority voting manifold alignment and object classification with KNN classifier. For the second case, an Supervoxel segmentation technology is used for keeping the continuity of segmentation results on temporal dimension. Fig.1 is the framework of the proposed multitemporal high resolution images classification method.

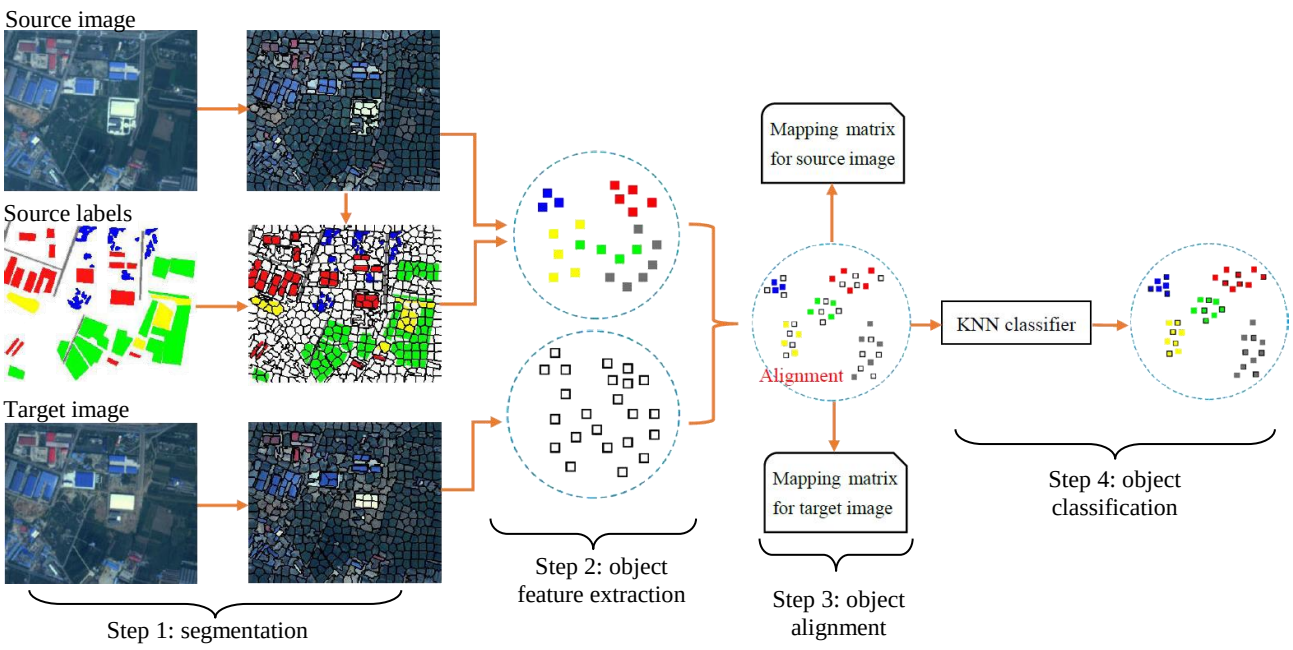


Fig.1 Framework of the proposed multitemporal high resolution images classification

2.1 Multi-temporal images segmentation

Two cases are considered in the segmentation parts: one is the all the image are from different areas (all images are segmented respectively) and another is that all images are from same area (all the images are segmented together into supervoxels).

2.1.1 Multi-temporal images segmentation for different areas condition

Simple linear iterative clustering (Achanta, 2012) is an adaption of *k*-means for super pixel generation and it is easy to use and be understand. By default, the only parameter to set in the algorithm is *k*, the desired number of approximately equally sized superpixels. For color images in CIELAB color space, the clustering procedure begins with an initialization step where some initial cluster centers are sampled on a regular grid spaced pixels apart. The clustering centers are moved to seed locations corresponding to the lowest gradient position in a S×3 neighborhood in order to avoid centering the edge of a superpixel, and to reduce the chance of seeding a superpixel with a noisy pixel. And then, in the assignment step, each pixel is associate with the nearest cluster center whose region overlaps its location in a size of S×S and 2S×2S dual window shown in Fig. 2. A distance measure *D*, which determines the nearest cluster center for each pixel, is then introduced. *D* computes the distance between each pixel and cluster center. The simply defining *D* will cause inconsistencies in clustering behavior for different superpixel sizes.

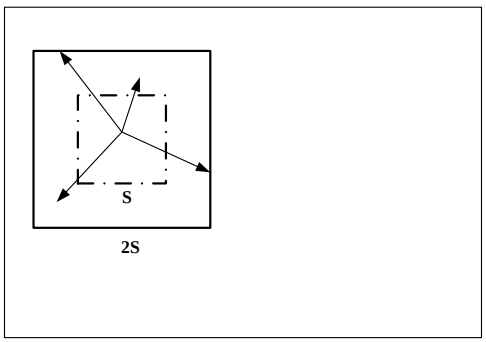


Fig.2. SLIC searches a limited region

In order to combine the color distance and spatial distance into a unified measure, it is necessary to normalize the color proximity and spatial proximity by their maximum distances within a cluster N_S and N_C, respectively. Doing so, *D'* is written as follows:

$$\begin{aligned} d_c &= \sqrt{\left(l_j - l_i\right)^2 + \left(a_j - a_i\right)^2 + \left(b_j - b_i\right)^2} \\ d_s &= \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2} \\ D' &= \sqrt{\left(\frac{d_c}{N_C}\right)^2 + \left(\frac{d_s}{N_S}\right)^2} \end{aligned} \tag{1}$$

The maximum spatial distance expected within a given cluster should correspond to the sampling interval, $N_S = S = \sqrt{(N/K)}$. Determining the maximum color distance N_C is not so straightforward, as color distances can vary significantly from cluster to cluster and image to image. This problem can be avoided by fixing N_C to a constant *m* so that (1) becomes

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \tag{2}$$

Once each pixel is connected with the nearest cluster center then the clusters centres for the mean [*l a b x y*] vector can be adjusted by an update step with all the pixels which belongs to the cluster. A residual error *E* is computed by the L_2 norm between the prior cluster center locations and the recent cluster center locations. The update and the assignment steps may be repeated till the error reduces.

2.1.2 Multi-temporal images segmentation for same area condition

In the 2.1.1 part, SLIC can also be extended to handle same area condition by including temporal dimension to the spatial proximity term of (1) as Eq.3. It's a supervoxel segmentation method. After supervoxel segmentation on the multi-temporal data, superpixel will be used in the next steps by separating the supervoxel to superpixel on the temporal dimension.

$$d_s = \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2 + \left(t_j - t_i\right)^2} \tag{3}$$

2.2 Object feature extraction and label reorganization

Because the spectral from same kind of objects are significantly different under the high spatial resolution and limited bands condition, averaging stage for getting the object features has the disadvantage for the object expression. For better express the object under the high resolution condition, an improved central spectrum used as the object feature in this paper. The central spectrum of each object can be calculated by using $\arg\max_{S_i} \sum_{j=1}^{k} \left\| S_i - S_j \right\|$ where S_i and S_j are pixels from the object and k is pixel number of the object. The label of object used in this paper is the label which has the largest number of categories in the statistics for all the labels on this object.

2.3 Majority voting manifold alignment for multi-temporal images alignment

Let $\mathbf{X} = \{x_i\}_{i=1}^{m} \in \mathbb{R}^p$ denote m labeled objects of source image and p is the feature bands. The class label of $\boldsymbol{X}$ is denoted as $C = \{c_i\}_{i=1}^{m}$. Let $\mathbf{Y} = \{y_i\}_{i=1}^{n} \in \mathbb{R}^p$ denote *n* un-labeled objects of target images. The aim of this letter is to learn mappings $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to map **X** and **Y** to a joint space **J**, where the manifold structures inside of **X**, **Y** and **J** will be preserved by using manifold technology.

2.3.1 Manifold alignment without corresponding

To preserve the manifold structures, some semi-supervised alignment [8], [12] directly computes the mapping results by minimizing the following cost function:

$$\begin{aligned} C(f,g) = \mu \sum_i \left(f_i - g_i\right)^2 + 0.5 \times \sum_{i,j} \left(f_i - f_j\right)^2 W_X^{i,j} \\ + 0.5 \times \sum_{i,j} \left(g_i - g_j\right)^2 W_Y^{i,j} \end{aligned} \tag{4}$$

where f_i is the mapping result of x_i, g_i is the mapping result of y_i and μ is the weight of the first term. W_X is the similarity matrix between each pixel from **X**, W_Y is the similarity matrix between each pixel from **Y**. The similarity between two pixels is calculated as following:

$$W_{a,b} = \exp\left(-\frac{\|\mathbf{a}-\mathbf{b}\|^2}{2\times\sigma_{spe}^2}\right) \tag{5}$$

The first term in (4) penalizes the differences between **X** and **Y** on the mapping results of the corresponding instances. The second and third terms in (4) ensure that the neighborhood relationship within **X** and **Y** will be preserved. Because the local geometry has more similarity than global geometry, only local neighborhood relationship is used to calculate the similarity matrix in most of manifold methods, such as locally linear embedding (LLE) [13] and Laplacian eigenmaps (LE). In other words, only the most similar pixel or top *k* similar pixels will be used to calculate the similarity matrix.

In multi-temporal HR images, the corresponding pixels are not easily to be found, so the first term in (4) should be changed to suit the condition without corresponding. To achieve this goal, all the data are calculated rather than only the corresponding pixels. Locality preserving projection, which has same goal with our expectation, can be are used to update the first term. Besides, because HR remote sensing images always have huge number of pixels, direct embedding methods have to consume a large amount of computation time to align all pixels of target image. Therefore, we seek for linear mapping functions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ ($\mathbf{X}\boldsymbol{\alpha}=\boldsymbol{f}$, $\mathbf{Y}\boldsymbol{\beta}=\boldsymbol{g}$) rather than direct embedding, so that the mapping can be used to project the rest data of target image into joint space directly (only a few of pixels of target image are used to compute the mapping matrix). After that, the mapped data of the rest of target image can be classified easily by using $\boldsymbol{f}$ (the mapped data of **X**) and its corresponding class label *L*. Based on those considerations, the cost function is modified as follows:

$$\begin{aligned} C(\alpha,\beta) &= \mu\sum_{i,j}\left(\alpha^T x_i-\beta^T y_j\right)^2 W_{XY}^{i,j} \\ &+\frac{1}{2}\left(\sum_{i,j}\left(\alpha^T x_i-\alpha^T x_j\right)^2 W_X^{i,j}+\sum_{i,j}\left(\beta^T y_i-\beta^T y_j\right)^2 W_Y^{i,j}\right) \end{aligned} \tag{6}$$

where W_{XY} is defined as the similarity matrix between each pixel of **X** and each pixel of **Y**. To align X and Y, we should find the solution to minimize the cost function $C(\alpha,\beta)$. In [13], the optimization problem can be written as follow:

$$\begin{aligned} &\min C(\alpha,\beta) \\ &= \min \gamma^T \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} L_X+\mu\Omega_X & -\mu W_{XY} \\ -\mu W_{XY}' & L_Y+\mu\Omega_Y \end{pmatrix}\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}^T \gamma \\ &\quad s.t.\ \gamma^T \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} D_X & 0 \\ 0 & D_Y \end{pmatrix}\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}^T \gamma = 1 \end{aligned} \tag{7}$$

where $\gamma=\left(\alpha^T,\beta^T\right)^T$, L_X and L_Y are the Laplacian matrix, $L_X=D_X-W_X$ and $L_Y=D_Y-W_Y$. D_X and D_Y are diagonal matrix, $D_X^{i,i}=\sum_j W_X^{i,j}$ and $D_Y^{i,i}=\sum_j W_Y^{i,j}$. Ω_X is an m×m diagonal matrix and $\Omega_X^{i,i}=\sum_j W_{XY}^{i,j}$, Ω_Y is an n×n diagonal matrix, and $\Omega_Y^{i,i}=\sum_i W_{XY}^{i,j}$.

By using the Lagrange trick, the solution to the optimization problem (7) is a typical generalized eigenvalue problem and can be solved as follow:

$$ZRZ^T\gamma=\lambda ZDZ^T\gamma \tag{8}$$

where Z, R and D is defined as follow:

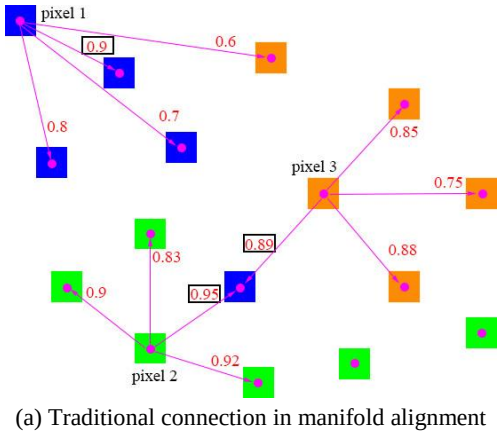
$$Z=\begin{pmatrix} X & \\ & Y \end{pmatrix} \tag{9}$$

$$R=\begin{pmatrix} L_X+\mu\Omega_X & -\mu W_{XY} \\ -\mu W_{XY}' & L_Y+\mu\Omega_Y \end{pmatrix} \tag{10}$$

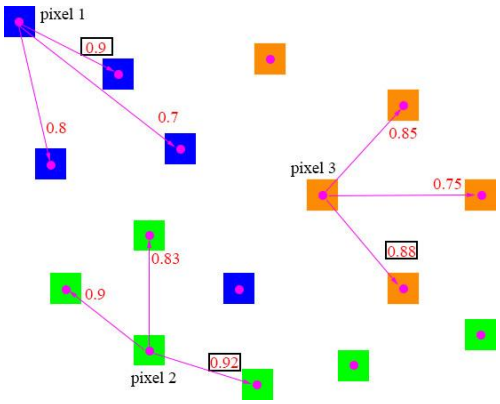
$$D=\begin{pmatrix} D_X & 0 \\ 0 & D_Y \end{pmatrix} \tag{11}$$

2.3.2 Majority voting stage in manifold alignment

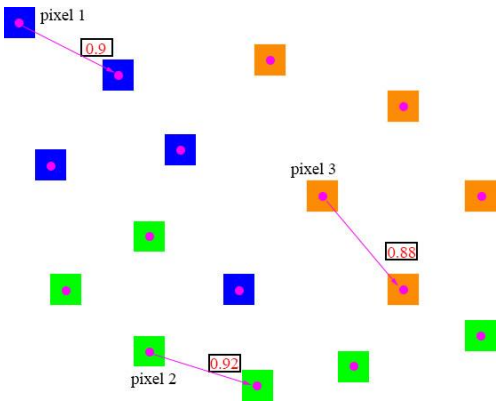
In manifold alignment methods, the most similar pixel or top k similar pixels are used to calculate similar matrix and Laplacian matrix for keeping local neighbourhood relationship. However, the most similar pixel or some of the top k similar pixels maybe not belong to same category with the current pixel. Fig.3a shows this case within source image (class label is known in source image). The most similar pixel of pixel2 (similarity is 0.95) is not belong to same category with the current pixel. One pixels of the top four similar pixels is not the same. In consequence, local neighbourhood relationship is preserved in manifold learning methods but inter-class distances are not improved. Fortunately, most of the similar pixels are still belong to same category with the current pixel (like in Fig.3a, four most similar pixels are used, three connections are current and one connection is wrong). In other words, the current pixel is more likely to belong to the category which the majority of pixels belong (shown in Fig.3b). If only the pixels from same category are used in manifold representation (connected pixels shown in Fig.3b), it will help us get rid of many false matches. And inter-class distances will be improved while keeping local geometry. For improving calculation efficiency, one of the effective ways is to reduce the connection number. The best connection (shown in Fig.3c) is the most similar pixel which has same category with current pixel (shown in Fig.3b).



(a) Traditional connection in manifold alignment



(b) Our connection

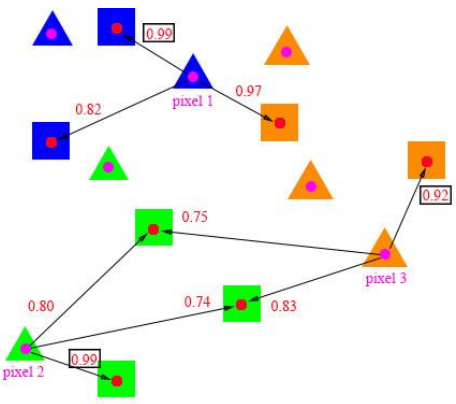


(c) Best single connection

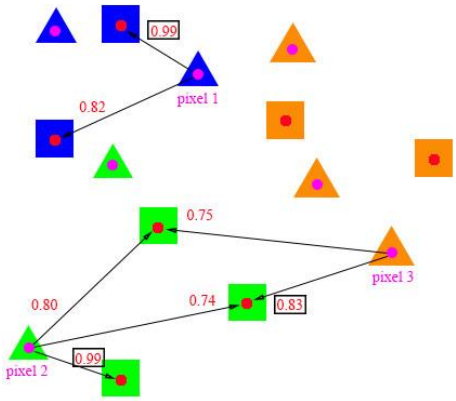
Fig.3 Majority voting strategy in source image

Fig.4a shows the same case (like Fig.3a) between source image and target image (current pixel is from target image and similar pixels are from source image). The most similar pixel or top k similar pixels maybe still not belong to the same category with the current pixel. Although the label of the current pixel is unknown, we still can assume that the current pixel is likely to belong to the category which the majority of pixels belong to. This can be achieved easily because the label of source image is known. It can be seen from Fig.4b that category of the maximum number of connections and all of its connection. Fig.4c is the best connection (maximum similarity in Fig.4b).

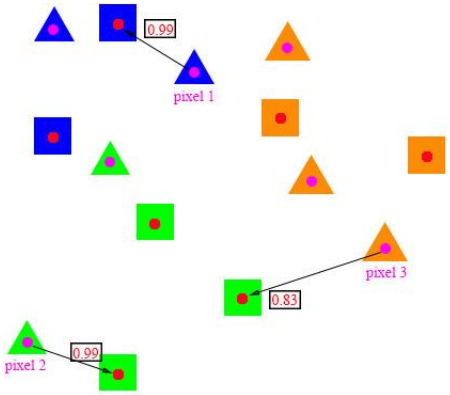
In this paper, the similar pixel used in majority voting MA is only the best single connection shown in Fig.3c and Fig.4c.



(a) Traditional connection in manifold alignment



(b) Our connection



(c) Best single connection

Fig 4. Majority voting strategy between source and target

2.4 Multi-temporal images classification

After all the mapping matrixes have been calculated (mapping matrix for objects of source images and mapping matrix for objects of target image), all the object from all domain/images can be mapping into the new alignment space. With some labelled object from source images, the objects from target image can be easily classified in alignment space by using KNN method. The unlabelled objects of source image can also be classified in alignment space with same method. If necessary, change detection will be achieved with the classification result of multi-temporal images.

3. EXPERIMENTS AND RESULTS

In this section, the proposed object-based manifold alignment for high resolution multi-temporal images classification method is tested. First, two groups of multi-temporal GF remote sensing scene data sets are introduced. Then, experimental results are given for the proposed method and the comparison methods.

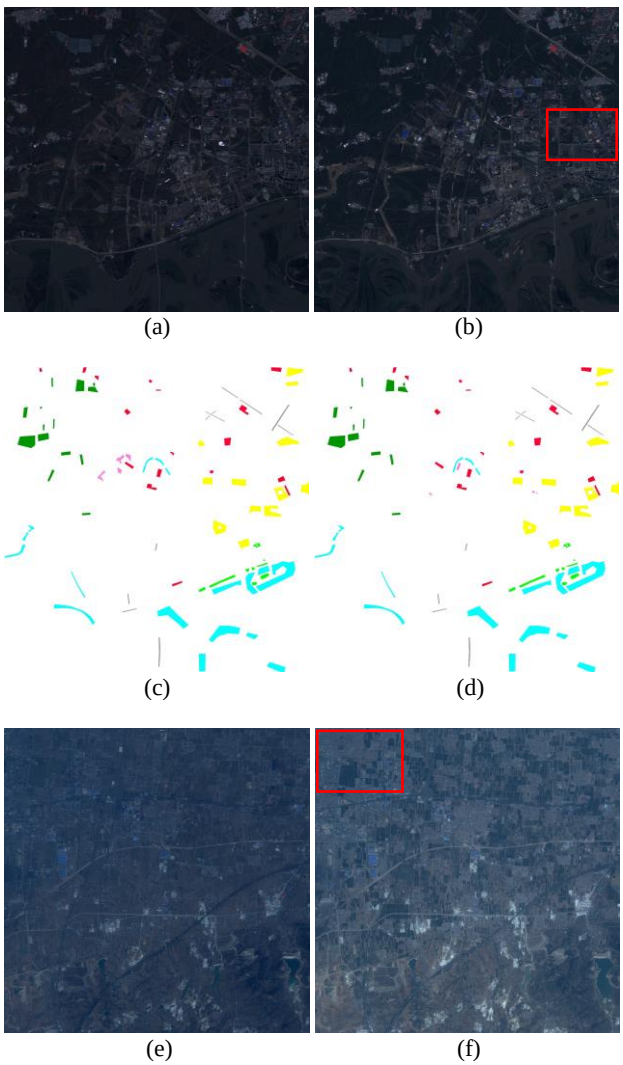
3.1 Multi-temporal HR remote sensing images

The proposed method was implemented on GF1 and GF2 data sets. Both GF1 and GF2 satellites are belong to China High-resolution Earth Observation System (CHEOS). CHREOS provides Near-Real-Time observations for disaster prevention and relief, climate change monitoring, geographical mapping, environment and resource surveying. CHEOS is composited by four subsystems: space-based system, near space and airborne system, ground system and application system. GF1 and GF2 are parts of space-based system. GF1 is configured with two 2m

panchromatic/8m multispectral camera and four 16 m multispectral medium-resolution and wide field camera set. GF2 has higher spatial resolution than GF1 and is capable of collecting satellite imagery of 0.8m panchromatic and 3.2m multispectral bands. Both GF1 and GF2 have four multispectral bands (blue, green, red and NIR) and have very similar band settings.

The GF1 data set used in this paper has two images (1536×1536) collected over Harbin city, Heilongjiang province, China on September 3, 2013 (set as source image, shown in Fig.5a) and September 20, 2014 (set as target image, shown in Fig.5b). The GF2 data set also has two images (2560×2560) collected over Qingdao city, Shandong province, China on March 2, 2014 (set as source image, shown in Fig.5e) and April 15, 2015 (set as target image, shown in Fig.5f). The two images of GF1 data set has small spectral drift but has big object changing. The two images of GF2 data set have few objects changing but has big spectral drift.

All images are from 8 typical scene categories, the class names and the color settings are shown in Fig.5h. Labelled data of each image is shown in Fig.5. All the images of each group are collected at same geographical area for utilizing the spatial information. All the images of each group must be registrated before being used.



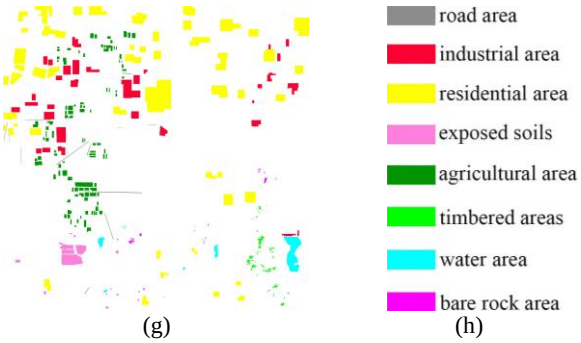


Fig.5 Multi-temporal HR GF data sets. Red–green–blue composite images of (a) source image of GF1, (c) target image of GF1, (e) source image of GF2, and (f) target image of GF2. Labeled data for (b) source image of GF1, (d) target image of GF1, and (g) both source image and target image of GF2. Class names of GF data sets (h).

3.2 Experiment settings

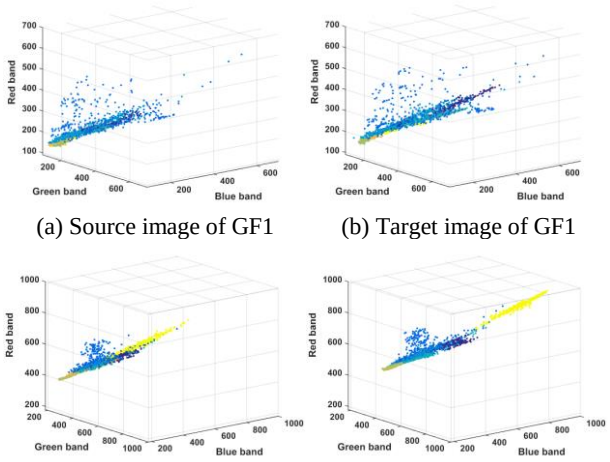
To verify the efficiency and superiority of the proposed objected-based multi-temporal high resolution classification method, three pixelwise methods are used to compare: direct classifier (using k-NN classifier without alignment), transfer component analysis (TCA, a state-of-the-art DA method) and LapSVM method (a typical kernel method in DA). TCA and LapSVM are compared because both of them have ability to compute the mapping matrix as well as the proposed method. Three aspects will be tested and analysed: classification performance on different data set (GF1 images and GF2 images), separability of data in alignment space, and the influence of the segmentation scale parameter. Classification accuracy will be calculated by running 20 times and take the mean to eradicate any discrepancies. Because the scenes of GF1 and GF2 data sets are so big that experiment result (only the result images) cannot display the details, only same small places of result images are used to be shown. The small areas used for showing are shown in Fig. 5b and Fig.5f (red box).

3.3 Experiment results and analysis

3.3.1 Alignment performance analysis

Fig. 6 is scatter plots of all the images in original space (only RGB bands) and each color corresponds to a class. 300 pixels per class are selected in each source image and target image for the scatter plots. It's easy to see that data from different class mix seriously and difficult to be classified. By comparing source and target image (Fig.6a and Fig.6b, Fig.6c and Fig.6d), same classes from two collection time are obviously different. That's why directly classify the target image using labeled source image always has low classification accuracy. In another words, temporal alignment before multi-temporal classification is very necessary.

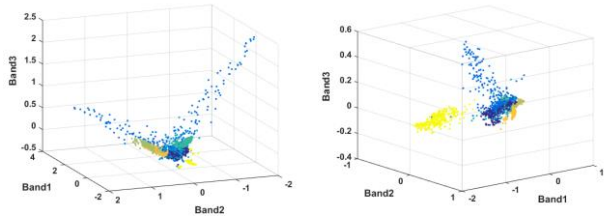
Fig.7 are the scatter plots of object feature from target images (both GF1 and GF2) in alignment space mapped by the proposed majority voting manifold alignment. The training samples used in Fig.7 are same with the samples used to plot Fig.6. In Fig.7 (in alignment space), the distances between different classes are enlarged, and the class separability is enhanced significantly. That mean the alignment parts of our proposed method (Majority voting manifold alignment) works well.



(a) Source image of GF1 (b) Target image of GF1

(c) Source image of GF2 (d) Target image of GF2

Fig.6 Scatter plots in original space (RGB bands) on 2 data sets.



(a) Target object feature GF1 (b) Target object feature GF2

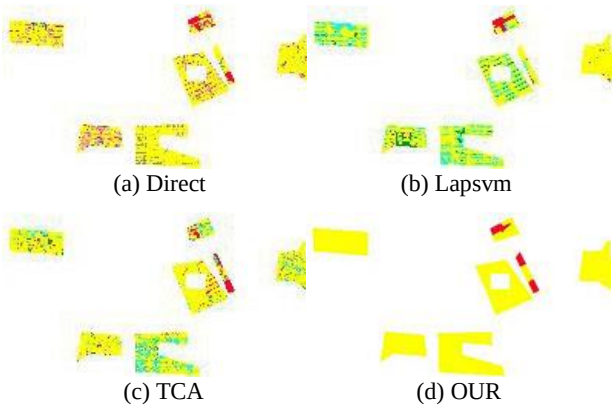
Fig.7 Scatter plots of the three first dimensions of target object feature in the alignment space on two data sets.

3.3.2 Classification performance

Table I is the classification accuracy of the two GF data sets. In the experiment, 100 samples of each class are used in the first three methods (direct classification, TCA and Lapsvm). For our object-based method, 10 objects of each class are used. In Table I, our method has the higher classification accuracy both on GF1 and GF2 data sets than the traditional methods. Fig.8 and Fig.9 are the classification maps on the small GF1 and GF2 data sets. Our object-based method has better result map and can overcome the "pepper and salt" problem on both data sets.

Table I Classification accuracy of the 2 GF data sets.

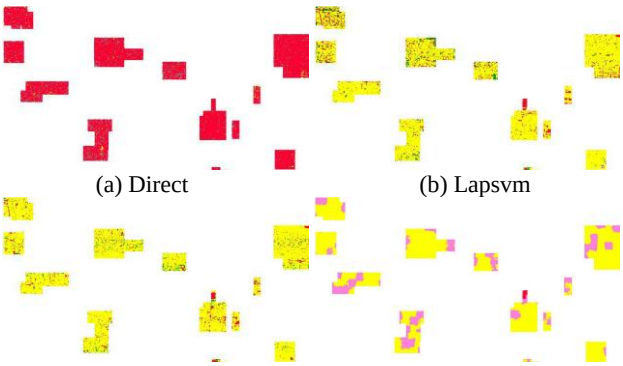
data	direct	TCA	Lapsvm	OUR
GF1	44.04	62.80	67.55	75.44
GF2	26.09	62.44	71.07	82.25



(a) Direct (b) Lapsvm

(c) TCA (d) OUR

Fig.8 Classification maps on the small GF1 data sets.



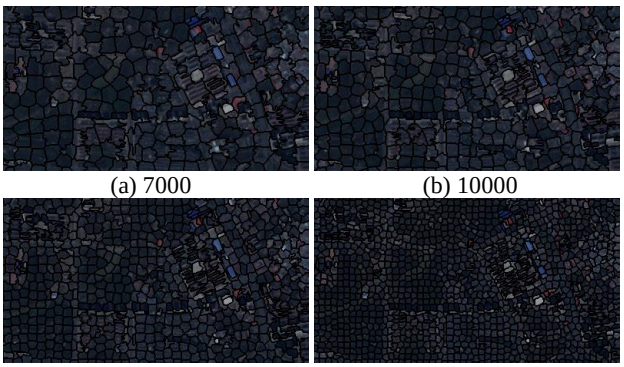
(a) Direct (b) Lapsvm

(c) TCA (d) OUR

Fig.9 Classification maps on the small GF2 data sets.

3.3.3 Segmentation scale impact analysis

As traditional object-based classification method, segmentation scale also has obviously effect on multitemporal high resolution classification. For testing relationship between the scale parameter and the classification performant, four segmentation scales (superpixels number for whole image, cut into 7000, 10000, 20000 and 40000 respectively) are tested on GF1 data set. Fig.10 is the segmented images of the using SLIC method. Table II is the classification accuracy on different segmentation scales. With the number of superpixel reducing, the accuracy increases gradually. Fig.11 is the corresponding classification maps to Fig.10 and Table II. "pepper and salt" problem has been improved better with small superpixels number.



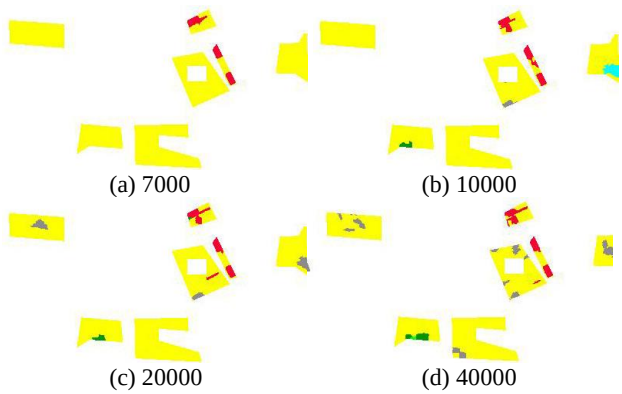
(a) 7000 (b) 10000

(c) 20000 (d) 40000

Fig.10 Segmentation result of different scales on small GF1 data.

Table II Classification accuracy of different segmentation scales.

data	7000	10000	20000	40000
GF1	82.25	79.04	78.63	74.89



(a) 7000 (b) 10000

(c) 20000 (d) 40000

Fig.11 Classification maps of different segmentation scales.

4. CONCLUSION

In this paper, an object-based multi-temporal high resolution images classification method is proposed. The main idea of proposed method is overcoming the "salt and pepper" problem and overlap problem in high resolution multitemporal images. SLIC segmentation is used and Majority voting manifold alignment is proposed for achieving this goal. Two groups of GF data sets are collected for evaluation. The experimental results validated the effectiveness of the proposed method, and the results also show that the proposed methods not only outperform the pixelwise multi-temporal images classification methods (TCA and LapSVM) obviously, but also effectively overcome the problem of "pepper and salt".

ACKNOWLEDGEMENTS

This work was supported by the National Science Fund for Excellent Young Scholars under the Grant 61522107 and the Natural Science Foundation of China under the Grant 61371180 and 60972144. We would like to thank the Heilongjiang Data and Application Center under the High Resolution Earth Observation System for providing GF1 and GF2 images.

REFERENCES

Tuia, D., Marcos, D., and Camps-Valls, G., 2016a. Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. ISPRS Journal of Photogrammetry and Remote Sensing 120, pp. 1-12.

Bruzzone, L., and Marconcini, M., 2009. Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy. IEEE Transactions on Geoscience and Remote Sensing 47(4), pp. 1108-1122.

Long, M., Wang, J., Sun, J., and Philip, S. Y., 2015. Domain invariant transfer kernel learning. IEEE Transactions on Knowledge and Data Engineering 27(6), pp. 1519-1532.

Kim, W., and Crawford, M. M., 2010. Adaptive classification for hyperspectral image data using manifold regularization kernel machines. IEEE Transactions on Geoscience and Remote Sensing 48(11), pp. 4110-4121.

Duan, L., Tsang, I. W., and Xu, D., 2012. Domain transfer multiple kernel learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(3), pp. 465-479.

Yang, X., Fu, H., Zha, H., and Barlow, J., 2006. Semi-supervised nonlinear dimensionality reduction. In: ACM International conference on Machine learning (ICML), pp. 1065-1072.

Wang, C., and Mahadevan, S., 2009. Manifold Alignment without Correspondence. In: International Joint Conferences on Artificial Intelligence Organization (IJCAI), pp. 1273-1278.

Yang, H. L., and Crawford, M. M., 2016. Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 54(1), pp. 51-64.

Tuia, D., Volpi, M., Trolliet, M., and Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 52(12), pp. 7708-7720.

Tuia, D., & Camps-Valls, G., 2016b. Kernel manifold alignment for domain adaptation. PloS one 11(2), pp. e0148655.

Tuia, D., Persello, C., and Bruzzone, L., 2016c. Domain adaptation for the classification of remote sensing data: an overview of recent advances. IEEE Geoscience & Remote Sensing Magazine 4(2), pp. 41-57.

Rasmussen, C., 2007. Superpixel analysis for object detection and tracking with application to UAV imagery. Advances in visual computing, pp. 46-55.

Mi, B., and Ko, J., 2009. Semantic segmentation of street scenes by superpixel co-occurrence and 3d geometry. In: IEEE International Conference on Computer Vision Workshops (ICCV), pp. 625-632.

Tong, N., Lu, H., Zhang, L., and Ruan, X., 2014. Saliency detection with multi-scale superpixels. IEEE Signal processing letters 21(9), pp. 1035-1039.

Liu, B., Hu, H., Wang, H., Wang, K., Liu, X., and Yu, W., 2013. Superpixel-based classification with an adaptive number of classes for polarimetric SAR images. IEEE Transactions on Geoscience and Remote Sensing 51(2), pp. 907-924.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence 34(11), pp. 2274-2282.