MARGINALIZED KERNEL-BASED FEATURE FUSION METHOD FOR VHR OBJECT CLASSIFICATION

Chuntian Liu, Wei Wei, Xiao Bai

School of Computer Science and Engineering
Beihang University
Haidian District, Beijing, China

Jun Zhou

Information and Communication Technology
Griffith University
Nathan, QLD 4111, Australia

ABSTRACT

Many image features can be extracted from very high resolution remote sensing images for object classification. Proper feature combination is a step towards better classification performance. In this paper, we propose a logistic regression-based feature fusion method which assigns different weights to different features. This method considers the probability that two images belongs to the same classes and the image-to-class similarity to define the similarity between two objects. This similarity is used as a marginalized kernel for the final classifier construction. Experiments on remote sensing images suggest that this approach is effective in various feature combination, and has outperformed the SVM baseline method.

Index Terms— Feature fusion, remote sensing image, kernel method, land cover classification

1. INTRODUCTION

In the past decades, many low-level feature extraction methods have been proposed and proved to be successful for image classification [1, 2, 3]. Naturally, these features and their combinations have been introduced into different remote sensing applications, such as object detection and land cover classification [4, 5, 6]. Comparing with single feature, multiple features provide abundant information on objects, especially those in very high resolution (VHR) images. However, how to combine the mass of information more effectively is still a big challenge.

One solution is feature fusion, which assigns different weights to different features, or combines several features for final decision. For example, Huang et al [7] combines two types of linear feature according to their spatial relationship. Cross validation is used to validate the consistency of two features. Li et al [8] fuse color and texture into one feature for object detection. They combine color histogram and the uniform local binary patterns using kernel principal component analysis. Then the maximum likelihood approach is used to select optimal feature set from the fused features.

Nevertheless, measuring the relevance of each individual feature to image classes is highly dependent on the images to be classified. To handle this problem, Zhang *et al.* [5] introduced a path alignment method to linearly combine multiple features in order to obtain a unified low-dimensional representation of these features. Tuia *et al.* [4] proposed a multiple kernel learning method to learn relevant weights of different features. However, both methods have not considered the dependency between different features and image classes.

To overcome this limitation, we propose a method that utilizes marginalized kernel for feature fusion. This method first extracts color and shape features, and uses the bag-of-words (BoW) method to convert them into vectors. For each class, a classifier is then trained on the concatenated feature vector using L1-logistic regression (LR) method [9] to obtain a sparse representation of discriminative visual words. The weight contribution in each visual words are used to construct a marginalized kernel[10, 11]. This kernel takes into account the probability that two images belongs to the same class and the image-to-class distance. Therefore, the marginalized kernel covers more complete description of data distribution. We show that the proposed method is effective, allowing better classification accuracy than the Support Vector Machine (SVM) classifier based on radial basis function kernel. To our knowledge, this is the first time that marginalized kernel is introduced to remote sensing community.

2. FEATURE FUSION AND IMAGE CLASSIFICATION

In this section, we first introduce the notation of the proposed method. Then we describe a linear regression model which is used to learn a set of weighting parameters for extracted multiple features. These weights are used to compute marginalized kernels, which are then used to train an SVM classifier. In the rest of the paper, we call it a logistic regression feature fusion (LRFF) method.

2.1. Notations and Definition

We commence from the following definition. Assume we have a training set $S = \{(x_i, y_i)\}_{i=1...m}$ of m labeled objects, where each x_i is an image patch containing object with a specific type. $y_i \in \{1, \dots, N\}$ is the label of corresponding object, and $Y = \{y_1, \dots, y_m\}$. For each image patch, n different features can be extracted. Using the bag-of-words method [12], each type of feature can be clustered to generate separate codewords. By assigning features to the closest codes words, an image patch is converted to a vector of concatenated histogram of codewords. If the size of visual dictionaries are $d_1, \dots, d_n$ for each type of feature, respectively, then the length of the final feature vector is $d = \sum_1^n d_j$.

2.2. Logistic regression

To fuse multiple features and determine the most relevant ones for classification, we adopt a logistic regression (LR) model. For each class, the model learning is treated in a binary classification setting, *i.e.*, $y_i \in \{+1, -1\}$, in which $y_i = +1$ means the object belongs to the class and $y_i = -1$ means it is not. The goal of this step is to learn a class-specific weight β, which will be used in the final kernel construction step. The objective function of LR is written as

$$\hat{\beta} = argmin_{\beta}(\sum_i \log\left(1 + \exp\left(-y_i \beta^T x_i\right)\right) + \lambda||\beta||_1) \tag{1}$$

where $\lambda > 0$ is the regularization parameter, and β is a weight vector. This model learns a parameter β for each class. A codeword with a high weight contributes significantly to discriminate positive and negative examples. Here, L1-regularization term is used to penalize all weights equally. It also prevents overfitting and limit the number of codewords selected for the classification step.

2.3. Marginalized kernel

From the LR model, both linear and nonlinear information of the model can be obtained. The linear information corresponds to the distance to the hyper-plane is reflected in the term $\beta_y^T x$, while the nonlinear information is the conditional probabilities $p(y|x)$ given by the LR model. Both information will be used in the marginalized kernel [10], which is defined as follows:

$$K(x, x') = \sum_y \sum_{y'} p(y|x) p(y'|x') K_z(z, z') \tag{2}$$

where $p(y|x)$ and $p(y'|x')$ are probabilities that x and x' belong to classes y and y', respectively. $K_z(z, z')$ is a joint kernel over the labeled samples $z = (x, y)$, which is defined by:

$$K_z(z, z') = Sim(y, y') \times \beta_y^T x \times \beta_{y'}^T x' \tag{3}$$

Here $Sim(y, y')$ is the similarity between y and y' ($0 \leq S(y, y') \leq 1$). The term $\beta_y^T x$ correspond to the image-to-class distance which has been proved to be effective in the nearest-neighbor based image classification [13]. If the hyperplanes of classes y and y' are close, two objects locate on the same side of the hyperplanes lead to a positive product $\beta_y^T x \times \beta_{y'}^T x$. Consequently, the kernel $K(x, x')$ will return a high similarity. For simplicity purpose, let $Sim(y, y') = 1$ where $y = y'$ and $Sim(y, y') = 0$ otherwise. Then final marginalized kernel becomes:

$$K(x, x') = \sum_{y \in Y} p(y|x, \beta_y) \times p(y|x', \beta_y) \times \beta_y^T x \times \beta_y^T x' \tag{4}$$

Thus, the SVM classifier H is given by

$$H(x') = sgn\left(\sum_{i=1}^{m} y_i \alpha_i K(x, x') + b\right) \tag{5}$$

The above steps are summarized in Algorithm 1.

Algorithm 1 LRFF Algorithm

```
Input: Training data $S = \{(x_i, y_i)\}_{i=1...m}$, where $y_i \in \{1, \dots, N\}$
Output: Marginalized Kernel-based Classifier $H$
for Each class $i = 1 \dots N$ do
    Divide training data into two sets, $y_k = +1$ if $y_k = i$ and $y_k = -1$ otherwise;
    Use LR model to learn parameters $\hat{\beta}_i$ for each class;
end for
Calculate the marginalized kernel $K(x, x')$;
Learn the SVM classifier $H$.
```

The proposed method leads to sparse occurrence of codewords for each class due to the use of L1-norm. Furthermore, different feature weights give the intuition that the significance of features vary with respect to the class of objects. Example weights learned for two object classes are shown in Fig. 1.

Once the marginalized kernel-based classifier is learnt, it can be used to classify unseen image samples. For each novel image sample, multiple image features are extracted and converted to vectors using the bag-of-words method, following the steps in the training stage. Then the SVM classifier with marginalized kernels are used for classifying each unseen sample.

3. EXPERIMENTS

The experiments were run on a Quickbird[14] image collected in Shanghai, China, in 2008. The size of this image is 2000*2000 pixels with spatial resolution of 0.6m. It contains

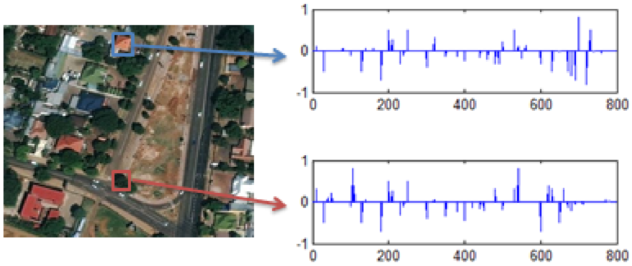

Fig. 1. Weight distribution on resident area and tree categories. The X axis represents length of image descriptor with multiple features. The Y axis shows the weight of each codewords, which has been normalized to $[-1, 1]$.

Table 2. Comparison of classification performance with different features

Method	Descriptor	Overall Acc.	Kappa
SVM (RBF)	*SIFT+Hue*	80.62%	76.88%
SVM (RBF)	*SIFT+LSS*	83.90%	79.65%
SVM (RBF)	*SIFT+Hue+LSS*	88.14%	84.07%
LRFF	*SIFT+Hue*	82.06%	78.08%
LRFF	*SIFT+LSS*	85.17%	81.34%
LRFF	*SIFT+Hue+LSS*	91.83%	86.23%

objects in five categories which correspond to five land cover types, *i.e.*, land, lawn, resident area (RA), road and tree. In the preprocessing stage, 1674 objects of these five classes have been segmented by a commercial software eCognition [15]. Then these objects were randomly split into a training set and a testing set. The experiments were run ten times with the average classification accuracy reported.

Table 1 lists three features used for the fusion and classification. Each feature is represented as a single vector. These feature vectors encode images properties, *i.e.*, shape and color related information. The number of codewords are set to be 300, 250 and 250 for SIFT, hue and LSS, respectively.

The proposed method was compared against a baseline SVM method which uses original feature with the radial basis function (RBF) kernel to construct the similarity matrices, $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \gamma > 0$. The penalty parameter C and kernel parameter γ are obtain using 5-fold cross-validations. Overall accuracy (OA) and Kappa coefficient are utilized to evaluate the classification performance.

The objective of the first experiment is to compare the performance of the proposed feature fusion method against the baseline method. The results are shown in Table 2. It can be observed from the table that fusion of three features is better than fusion of only two features. The LRFF has demonstrated clear advantage over the baseline method in both evaluation criteria. Fig. 2 shows results by the proposed method on an image patch cropped from the original image, with land cover types labeled in different colors.

To analyze the influence of the training set size to the classification accuracy, we have trained the classification model with different numbers of training samples. These included 50, 100, 150, 200 and 250 training samples for each class, respectively while the rest of the samples were put into the testing set. The curves in Figure 3 shows that the performance of the LRFF method and the baseline method improves with the increase of the number of training samples. The LRFF method has consistently performed better than the baseline SVM method.

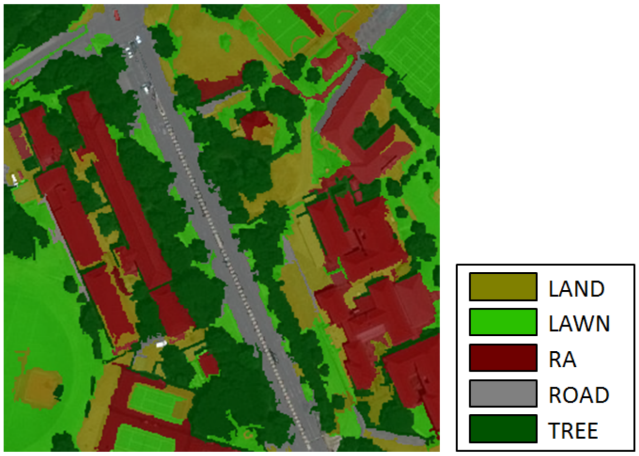

Fig. 2. Sample classification results by the proposed method.

4. CONCLUSION AND FUTURE WORK

In this paper, we have introduced a method to combine different features for object classification. The success of the proposed method is due to the following two reasons: 1) the learned conditional probabilities of the LR model are used in the kernel construction, 2) image-to-class similarity has been taken into account to define the similarity between two objects. Our future work will focus on more advanced feature combination method in order to construct more powerful kernel description in image classification.

5. ACKNOWLEDGMENT

This research is supported under the National Natural Science Foundation of China projects No. 61105002.

Table 1. Feature descriptors used in this paper

Feature	Description
SIFT[1]	A scale invariant feature transform descriptor
Hue[2]	Local features with color information
LSS[3]	A descriptor integrated with the contextual and shape information

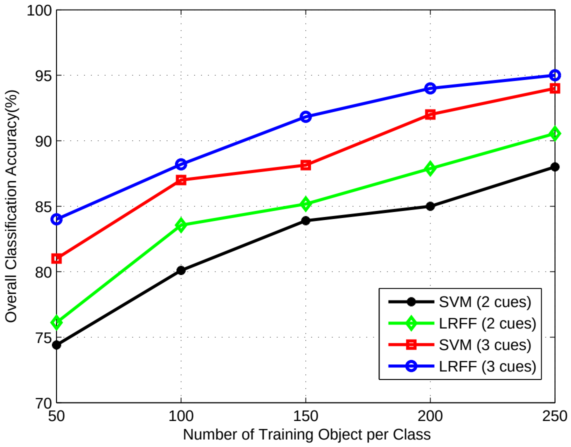

Fig. 3. Influence of number of training samples.

6. REFERENCES

[1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] J. Van De Weijer and C. Schmid, "Coloring local feature extraction," *Proceedings of the ECCV*, pp. 334–348, 2006.

[3] Eli Shechtman and Michal Irani, "Matching local self-similarities across images and videos," in *Proceedings of the CVPR*, 2007, pp. 1–8.

[4] Devis Tuia, Gustavo Camps-Valls, Giona Matasci, and Mikhail Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3780–3791, 2010.

[5] Lefei Zhang, Liangpei Zhang, Dacheng Tao, and Xin Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 879–893, 2012.

[6] Xin Huang, Liangpei Zhang, and Pingxiang Li, "An adaptive multiscale information fusion approach for feature extraction and classification of $ikonos$ multispectral imagery over urban areas," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 654–658, 2007.

[7] Z. Huang, J. Zhang, L. Wang, and F. Xu, "A feature fusion method for road line extraction from remote sensing image," in *Proceedings of the IGARSS*, 2012, pp. 52–55.

[8] Z. Li, Y. Liu, R. Hayward, and R. Walker, "Color and texture feature fusion using kernel pca with application to object-based vegetation species classification," in *Proceedings of the ICIP*, 2010, pp. 2701 – 2704.

[9] M. Collins, R.E. Schapire, and Y. Singer, "Logistic regression, adaboost and bregman distances," *Machine Learning*, vol. 48, no. 1, pp. 253–285, 2002.

[10] H. Kashima, K. Tsuda, and A. Inokuchi, "Marginalized kernels between labeled graphs," in *Proceedings of the ICME*, 2003, pp. 321–328.

[11] Koji Tsuda, Taishin Kin, and Kiyoshi Asai, "Marginalized kernels for biological sequences," *Bioinformatics*, vol. 18, no. suppl 1, pp. S268–S275, 2002.

[12] Sheng Xu, Tao Fang, Deren Li, and Shiwei Wang, "Object classification of aerial images with bag-of-visual words," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 2, pp. 366–370, 2010.

[13] E.Shechtman O.Boiman and M.Irani, "In defense of nearest-neighbor based image classification," in *Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[14] D. Stow, A. Lopez, C. Lippitt, S. Hinton, and J. Weeks, "Object-based classification of residential land use within accra, ghana based on quickbird satellite data," *International Journal of Remote Sensing*, vol. 28, no. 22, pp. 5167–5173, 2007.

[15] Ursula C Benz, Peter Hofmann, Gregor Willhauck, Iris Lingenfelder, and Markus Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3, pp. 239–258, 2004.