SEGMENTATION OF MULTISOURCE IMAGE BY MULTIVIEW LEARNING

Keming Chen[1], Jian Cheng[1], Hanqing Lu[1], Zhixin Zhou[12]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190;
[2]Beijing Institute of Remote Sensing, Beijing, China, 100854
E-mail: {kmchen, jcheng, luhq@nlpr.ia.ac.cn}, zhixin.zhou@ia.ac.cn

1. INTRODUCTION

Image segmentation is a critical technique for remote sensing image analysis and interpretation. It is the basis of image understanding, such as region-based change detection for maps updating, target recognition, and so on. Existing literatures [1, 2] have proved that, by combining multiple data sources, e.g., synthetic aperture radar (SAR) and optical imagery, the overall classification accuracy will significantly increase compared to the quality of a single-source classifier.

Multi-view learning approaches [3, 4, 5] form a class of semi-supervised learning techniques using multiple views to effectively learn from partially labeled data. Although there are a variety of multi-view learning algorithms, they are all founded on the common principle: training the classifiers for the views by maximizing the agreement on their predictions for the unlabeled data， $min \sum_{x_u \in U} \sum_{i \neq j} \| f_i(x_u^i) - f_j(x_u^j) \|^2$， where f_i， $i = 1, 2$ is the prediction from the classifier of *i-th* view for the unlabeled data point $x \in U$. For remote sensing image processing, the views can be formed from the same modality (e.g., images acquired in the same geographical area at two different times by the same sensor) or multiple sources (e.g., images acquired in the same geographical area by different sensors). In the aids of multiple redundant views, existing approaches to multi-view learning or co-training [3] achieve a more robust result by working in a bootstrap mode and mutually training classifiers to augment the labeled data. Multi-view learning has shown its advantages over single view learning, especially when the theoretical guarantees, i.e. compatible and redundancy, are satisfied [3, 5].

In this paper, a novel method for the joint segmentation of multisource images is formed as a multiview learning problem. Each individual data source, i.e., the SAR data or the optical image, is assigned to be an individual view in multiview learning. Markov random field (MRF) model is employed for each view segmentation and energy function is defined on every view. Therefore, the joint segmentation of multisource data sets is transformed into simultaneously minimizing the combined MRF energy functions defined on individual view. This gives a new perspective on the previously multisource imagery segmentation methods, and also makes explicit the circumstances under which these algorithms fail.

2. MULTISOURCE IMAGE SEGMENTATION BASED ON MULTIVIEW LEARNING

Structure of the proposed model is illustrated in *Fig. 1*.

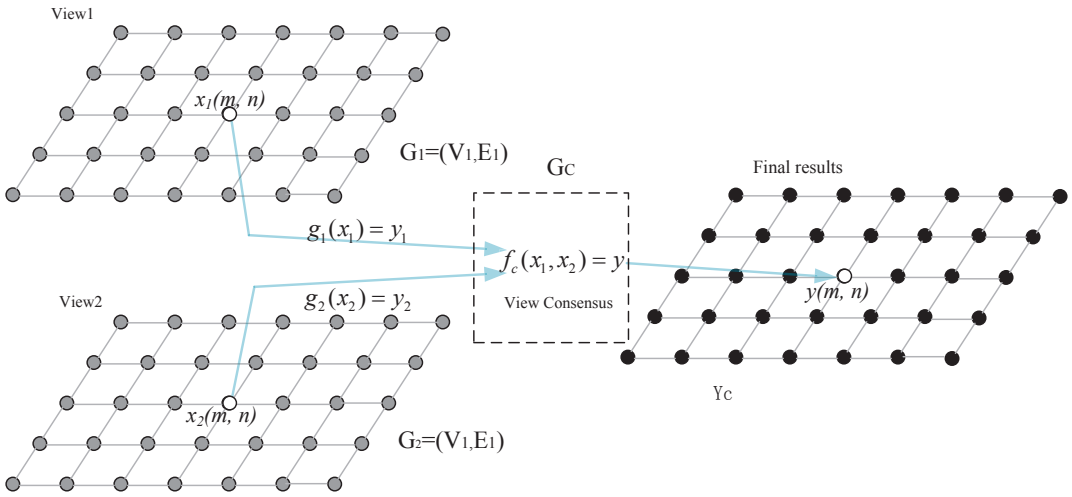


Figure 1. Structure of multisource image segmentation model based on multiview learning.

Details of the proposed approach are described as following:

In the proposed approach, the SAR image X_1 and the optical image X_2 are considered as two views in the multiview learning. Let $X_i=\{ x_i(m, n), 1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N\}$ $(i=1,2)$ and $x_i(m, n)$ be the features in location (m, n) obtained from the *i-th* view. Then the vector X_i represents all sample observations for the *i-th* view. $Y_i=\{ y_i(m, n), 1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N \}$ $(i=1,2)$ is the segmented map of the *i-th* view. Look at the multi-view learning from Markov random field model perspective, each view is modeled by a MRF model and an undirected graph is defined on each data set. *Let* $G_i=(V_i, E_i)$ denote the graph model defined on the *i-th* view, and E_i is the corresponding Gibbs energy function. Hence, we denote $g_i(x_i(m,n)) = y_i(m,n)$ for each view. Although there are multiple views from the feature sets, one pixel in the final change map can only has one label $y(i, j)$. Here we define an additional combined function G_c and denote Y_c as the final consensus output of the multiple views. To model the dependency among different views and reach a consensus, we introduce a latent function F. So the global observation process is expressed by $F(x) = \{f_i(x_i)\}$. Considering image labeling as an inverse process that attempts to estimate the best Y given the observed image X. It can be formulated as a maximum a posterior (MAP) problem for which maximizing the posterior $P(Y|F)$ gives a solution. According to Bayes rule, this is equivalent to maximizing $P(y|x)P(x)$. The most probable or MAP labeling x is defined as:

$$y = \arg\max_{y \in Y} \{p(F(x) \mid y) p(y)\} . \qquad (1)$$

Suppose the different nodes in multi-views be conditionally independent given a consensus label. Thus, P(F|y) can be obtained by as a product of singleton probability terms assigned to the nodes in each view:

$$P(F(x)\mid y)=\prod_i P(f_i(x_i)\mid y_i)$$
$$=\prod_i\sum_{m=1}^{M}\sum_{n=1}^{N}\psi_i^1(x_i(m,n))+\psi_i^2(y_i(m,n),y_i(N_{mn})) \quad .(2)$$

Here for each view $\psi_i^1(x_i(m,n))$ specifies the feature model energy about the correlation of feature levels between the individual label y_i in final classification map and the pixel $x_i(m,n)$ in the observed image. $\psi_i^2(y_i(m,n),y_i(N_{mn}))$ describes the potential function of the interactions among pixels $y_i(m,n)$ in the appropriate neighborhoods N_{mn} defined on *i-th* view. Commonly, two models are popularly adopted for representing the feature model ($P(x|y)$) and the spatial context model ($p(y)$) in pairwise MRF model. For $P(x|y)$, we often consider x_i as a constant feature level corrupted by additive independent noise. Furthermore, it is often assumed to be a Gaussian function for simplicity,

$$\mathrm{P}(\mathrm{x_i}\mid \mathrm{y_i})=\frac{1}{\sqrt{2\pi\Sigma_i}}\exp\left[-\frac{1}{2}(x_i-\mu_i)^T\Sigma_i^{-1}(x_i-\mu_i)\right] . \qquad (3)$$

For $P(y)$, an MRF model named the multilevel logistic model (MLL) has been popular. Under the Markovian framework, the joint probability distribution of the image sites can be rewritten in terms of the local spatial interactions, which are analytically expressed by clique energy functions. The clique energy of MLL is defined as

$$P(y_i)=\frac{1}{Z}\exp\left[-\sum_{N_{mn}}\beta\delta(y_i,y(N_{mn}))\right] , \qquad (4)$$

where β is a penalty for the existence of boundary site pairs and Z is a normalizing constant known as the partition function. Substituting the feature model energy function (3) and prior potential functions (4) into (1), the MAP formulation of the labeling task (1) is transformed into an energy minimizing problem as

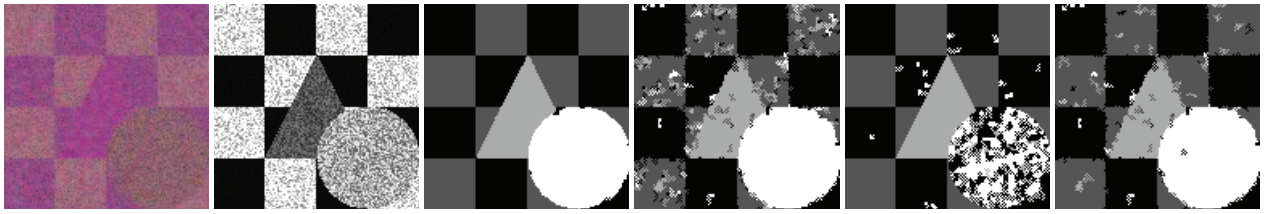
$$\arg\min\sum_{i=1,2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\{\frac{1}{2}\ln(2\pi\Sigma_i)+\frac{(x_i-\mu_i)^2}{2\Sigma_i}+\beta\sum_{N_{mn}}\delta(y_i,y(N_{mn}))\right\} .(6)$$

The energy function can be minimized by iterative conditional mode (ICM) [6] or simulated annealing (SA) [7].

3. EXPERIMENTS

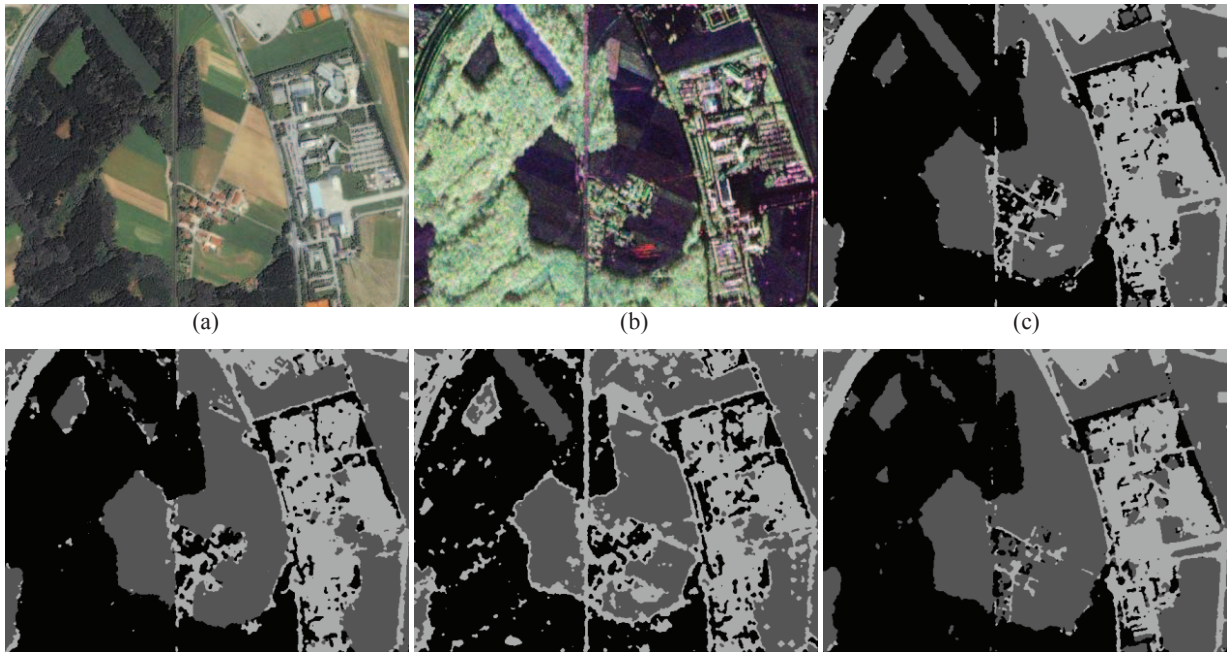
To thoroughly assess the performance of the proposed approach, we verify it on several artificial images and real remote sensing images. For synthetic data, here we only used intensity values of the image as the features. For real remote sensing images, we used *L*a*b* color feature and *Gabor* feature to describe the images. We also compared our results with ones obtained by Waske's method [1]. Some segmentation results are shown in *Fig. 2*

and *Fig. 3* for visual comparison. Experiments on synthetic dataset and real remote sensing images confirm the effectiveness of the proposed approach.



(a) (b) (c) (d) (e) (f)

Figure 2. Visual comparison of some segmentation results on synthetic data. (a) Optical image; (b) SAR image; (c) Segmentation results obtained by the proposed approach; and Segmentations on (d) optical image; (e) SAR image; (f) Waske's result.



(a) (b) (c)

(d) (e) (f)

Figure 3. Visual comparison of some segmentation results on real remote sensing datasets. (a) Optical image; (b) SAR image; (c) Segmentation results obtained by the proposed approach; and Segmentations on (d) optical image; (e) SAR image ; (f) Waske's result.

REFERENCES

[1] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[2] B. Waske and S. van der Linden, "Classifying multilevel imagery from SAR and optical sensors by decision fusion," IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pp. 1457–1466, May 2008.

[3] A. Blum and Tom Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," *Proc. Conference on Computational Learning Theory*, pp.92-100, July, 1998.

[4] S. Yu, B. Krishnapuram, R. Rosales, H. Steck, and R. B. Rao, "Bayesian co-training", *In Advances in Neural Information Processing Systems.* vol. 20, pp. 1665-1672, 2008.

[5] K. Chen, Z. Li, J. Cheng, Z. Zhou and H. Lu, "A Variational Co-training Framework for Remote Sensing Image Segmentation", *Int'l Geoscience & Remote Sensing Symposium, IEEE*, to be appeared, Jul. 2009.

[6] J. Besag. "On the statistical analysis of dirty pictures (with discussion)", *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[7] S. Barnard. "Stochastic stereo matching over scale", *International Journal of Computer Vision*, vol. 3, no. 1, pp. 17–32, 1989.