International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,
Volume XL-1/W1, ISPRS Hannover Workshop 2013, 21 – 24 May 2013, Hannover, Germany

MULTI-SOURCE HIERARCHICAL CONDITIONAL RANDOM FIELD MODEL FOR FEATURE FUSION OF REMOTE SENSING IMAGES AND LIDAR DATA

Z. Zhang [a], M.Y. Yang [b], M. Zhou[a]

[a] Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing, China
(zhangzheng, zhoumei)@aoe.ac.cn
[b] Institute for Information Processing (TNT), Leibniz University Hannover, Germany
yang@tnt.uni-hannover.de

Commission WG VI/4, WG III/3

KEY WORDS: Feature fusion, Conditional Random Field, Image Classification, Multi-source Data, Hierarchical model

ABSTRACT:

Feature fusion of remote sensing images and LiDAR points cloud data, which have strong complementarity, can effectively play the advantages of multi-class features to provide more reliable information support for the remote sensing applications, such as object classification and recognition. In this paper, we introduce a novel multi-source hierarchical conditional random field (MSHCRF) model to fuse features extracted from remote sensing images and LiDAR data for image classification. Firstly, typical features are selected to obtain the interest regions from multi-source data, then MSHCRF model is constructed to exploit up the features, category compatibility of images and the category consistency of multi-source data based on the regions, and the outputs of the model represents the optimal results of the image classification. Competitive results demonstrate the precision and robustness of the proposed method.

1. INTRODUCTION

Nowadays, there are many different sources of earth observation data which reflect the different characteristics of targets on the ground, so how to fuse the multi-source data reasonably and effectively for the application, such as object classification and recognition, is a hot topic in the field of remote sensing applications. In all the data mentioned above, remote sensing images and LiDAR points cloud have strong complementarity, so fusion of the two sources of data for object classification is attached more and more attention and many methods were proposed. In general they can be classified into image fusion (Parmehr et al., 2012; Ge et al., 2012) and feature fusion (Deng et al., 2012; Huang et al., 2011). The methods for image fusion always include different resolution data sampling and registration, so the processing is time-consuming, and will inevitably lose a lot of useful information, which reduces the accuracy of the subsequent image classification. In the feature fusion methods, the features are usually extracted independently from different sources data, and the fusion lacks consideration of correspondence of location and contextual information, so the classification results could be improved. In addition, because the features selected in some methods are not invariant to rotation, scale, or affine, they are always poor in stability. In order to overcome the shortages of former methods, this paper presents a novel multi-source hierarchical conditional random field (MSHCRF) model to fuse features extracted from remote sensing images and LiDAR data for image classification. Firstly, typical features are selected to obtain the interest regions from multi-source data. Then MSHCRF model is constructed to exploit up the features, category compatibility of images and the category consistency of multi-source data based on the regions, and the outputs of the model represents the optimal results of the image classification.

2. DESCRIPTION OF FEATURES SELECTED

In remote sensing images and LiDAR data, while the abundance of information offers more detailed information of interest objects, it also enhances the noises. Selection of appropriate features in a reasonable way is important in our method.

In order to provide a reliable basis for subsequent processing, the proposed model contains five kinds of typical features: local saliency feature (LSF), line feature (LF) and texture feature (TF) are extracted from remote sensing images, mean shift feature (MSF) and alpha shape feature (ASF) are from LiDAR data, so it's robust to background interference, change of scale and perspective, etc.

The detector of K&B (Kadir et al., 2001) is a representative LSF, which is invariant to viewpoint change, and sensitive to image perturbations. We utilize the detector of K&B to calculate saliency of each pixel in the images.

LSD is a linear-time line segment detector that gives accurate results, a controlled number of false detections, and requires no parameter tuning. In accordance with the method introduced in (Grompone et al., 2010), we can calculate the response value at each pixel.

As the basic unit of TF, Texton is utilized to distinguish between foreground and background regions effectively and increase the accuracy of the results. Similar to the method in (Shotton et al., 2009), we can obtain the response to Texton of each pixel in the image.

For the sparseness and discreteness of LiDAR points cloud data, we utilize an adaptive mean shift algorithm which is a sample

point estimation method based on data-driven. In our model, the specific process of achieving the MSF is introduced in (Georgescu et al., 2003).

Based on the planar features obtained, the alpha shape algorithm is used to extract the boundary contour of each target, and then the Delaunay triangulation is used to get the line feature of LiDAR points cloud. The extraction of the ASF refers to (Shen et al., 2011).

3. FEATURE FUSION USING MSHCRF

In the field of image processing, the regions of interest are usually detected independently, but considering the relative position between regions in single data and the correspondence between regions from multi-source data, the labelling of every region should not be independent. The Conditional Random Field (CRF) model is an effective way to solve the problem of prediction of the non-independent labelling for multiple outputs, and in this model, all the features can be normalized globally to obtain the global optimal solution.

In view of the advantages above, based on the standard CRF model, we propose the MSHCRF model to learn the conditional distribution over the class labelling given an image and corresponding LiDAR data, and the model allows us to incorporate LSF, LF, TF, MSF, ASF and correspondence information in a single unified model. The conditional probability of the class labels **c** given an image **I** and LiDAR data **L** is defined as follow

$$\log P(\mathbf{c}\mid\mathbf{I},\mathbf{L},\theta)=\sum_{i}P_1(c,x_i)+\sum_{(i,j)\in N}P_2(c,x_i,x_j)+\sum_{(i,k)\in H}P_3(c,x_i,l_k)-\log Z(\theta,\mathbf{I},\mathbf{L}) \quad (1)$$

where θ is the model parameters, $Z(\theta,\mathbf{I},\mathbf{L})$ is the partition function, i and j index nodes in the grid corresponding to the positions in the image, and k index nodes in the grid corresponding to the positions in the LiDAR points cloud. N is the set of pairs collecting neighborhood in the image and H is the set of corresponding pairs collecting neighborhood in both images and LiDAR data. P_1 is the unary potentials, which represent relationships between variables and the observed data. P_2 is the pairwise potentials, representing relationships between variables of neighboring nodes. P_3 is the hierarchical pairwise potential, which represents corresponding relationships between images and LiDAR data The full graphical model is illustrated in Figure 1.

3.1 Unary potentials

The unary potentials are consisted of three element: LSF, LF and TF potentials, predict the label x_i based on the image **I**

$$P_1(c,x_i)=LSF(c,x_i;\theta_{LSF})+LF(c,\mathbf{x};\theta_{LF})+TF(c,\mathbf{x};\theta_{TF}) \quad (2)$$

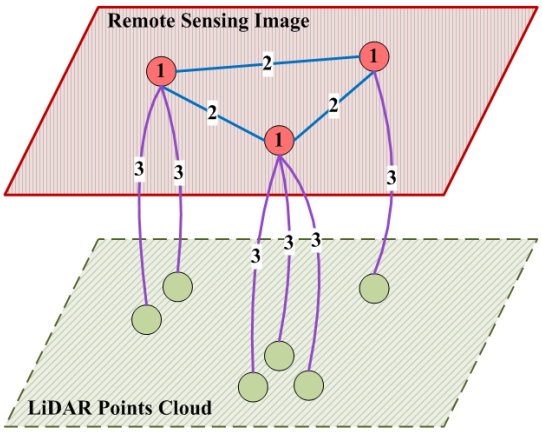


Figure 1. Illustration of the MSHCRF model architecture. Red nodes with No.1 correspond to regions extracted by the features selected in images, blue lines linking red nodes with No.2 represent the dependence between neighbor regions, and purple lines linking red and green nodes with No.3 indicate the hierarchical relation between regions from multi-source data.

In accordance with the methods described previously, we can calculate the *LSF*(i) of each pixel in the image to obtain the local saliency feature image, in which local saliency feature models are represented as mixtures of Gaussians (GMM), so there is

$$LSF(c,x_i;\theta_{LSF})=\log\sum_{k}\theta_{LSF}(c,k)G(Sa\mid\mu_k,\Sigma_k) \quad (3)$$

where k represents the component the pixel is assigned to, μ_k and Σ_k are the mixture mean and variance respectively, and parameter $\theta_{LSF}=\left(\sum_i\delta(c_i=c_i^*)P(k\mid x_i)+0.1\right)^3/\left(\sum_i P(k\mid x_i)+0.1\right)^3$ represents the distribution $p(c|k)$, the mixture term $p(k|x_i)\propto p(x_i|k)$, a class labeling c_i* is inferred, and $\delta(\cdot)$ is a 0-1 indicator function.

Similar to the LSF potentials, we can get the line segment image *LFI*(i) by calculating the LSD in the image. The LF potentials take the form of a look-up table

$$LF(c,\mathbf{x};\theta_{LF})=\log\theta_{LF}(c,\mathbf{x}) \quad (4)$$

where parameter $\theta_{LF}(c,\mathbf{x})=1-\left|\delta(c_i)-\delta(x_i)\right|-0.1\Big|$ represents the correlation between the value in *LFI*(i) and the label c.

Based on the boosting learning algorithm, we can obtain the classifier of Texton, to which the responses are used directly as a potential, so that

$$TF(c,\mathbf{x};\theta_{LF})=\log P(c\mid\mathbf{x},i) \quad (5)$$

where $P(c|\mathbf{x},i)$ is the normalized distribution given by the classifier using the learned parameters θ_{TF}.

3.2 Pairwise potentials

The pairwise potentials describe category compatibility between neighboring pixels x_i and x_j of the line segment image *LFI*(*i*) and the responses of Texton classifier on the image **I**. The pairwise potentials have the form introduced in (Yang et al., 2011), and the pairwise potentials are the sum of two kinds of responses.

3.3 Hierarchical pairwise potentials

Compared to the remote sensing images, LiDAR points cloud have the characteristics of sparseness and discreteness, which like the low-resolution images sampled from the corresponding images, and the features extracted from multi-source data are different. So in order to enhance the fusion performance, we introduce the hierarchical pairwise potentials, which represent correspondence between the multi-source data, in our MSHCRF model.

The hierarchical pairwise potentials describe category consisteny between the corresponding regions in multi-source data, from which we can obtain linear features, such as LF and ASF, and planar features, such as TF and MSF. In order to enhance the fusion performance, we refer to the consistency of the linear and planar features separately, note as $Diff_l(c,x_i,l_k)$ and $Diff_p(c,x_i,l_k)$. So there is

$$P_3(c,x_i,l_k) = Diff_l(c,x_i,l_k,\theta_l) + Diff_p(c,x_i,l_k,\theta_p) \quad (6)$$

For describing the consistency of linear features, we firstly normalize each value of TF and ASF to get the $\hat{x}_i$ and $\hat{l}_k$, then

$$Diff_l(c,x_i,l_k,\theta_l) = \theta_l \exp(-\varepsilon \mid \hat{x}_i - \hat{l}_k \mid^2)\delta(c(x_i) \neq c(l_k)) \quad (7)$$

where the comparative item $\varepsilon = (2<|\hat{x}_i - \hat{l}_k|^2>)^{-1}$, $<\cdot>$ indicates the global average, and θ_l needs to be selected manually to minimize the error on the validation set.

As to the consistency of planar features, the calculation is similar to the one of linear features.

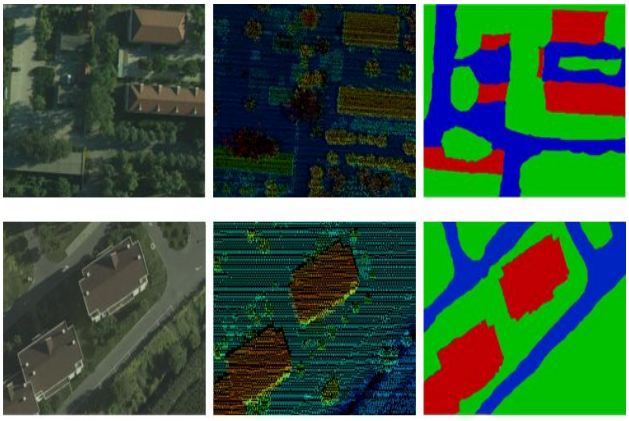
3.4 Image classification with the MSHCRF model

By the formula derivation, empirical deduction and training on validation data, each parameter can be learned respectively. Given a set of parameters learned for the MSHCRF model, the optimal labelling, which maximizes the conditional probability, is found by applying the alpha-expansion graph-cut algorithm (Boykov and Jolly, 2001), and it represents the optimal results of the image classification.

4. EXPERIMENTS

We conduct experiments to evaluate the performance of the MSHCRF model on the airborne data collected at Beijing, China, which include the remote sensing images with the resolution 1m and LiDAR points cloud with the point density 4 points/m^2. The objects in all images are labeled with 3 classes: building, road and trees. These classes are typical objects appearing in the airborne images. In the experiments, we take the ground-truth label of a region to be the majority vote of the ground-truth pixel labels, and randomly divide the images into a training set with 50 images and a testing set with 50 images.

Figure 2 shows the classification result from MSHCRF model. We run the experiment on the whole test set, and get the overall classification accuracy 73.6%. For comparison, we also carry out another experiment by removing the hierarchical pairwise potentials from the model, namely classifying only with the remote sensing images, which is similar to the standard CRF model (Shotton et al., 2009), and the overall accuracy is decreased to 68.9%. Therefore, the MSHCRF model increases the accuracy by 4.7%. The parameter settings, learned by cross validation on the training data, are θ_{TF} = 0.35, θ_l = 0.12, and θ_p = 0.15.



Remote sensing image LiDAR points cloud Classification result

Figure 2. The classification result from the MSHCRF model. (In the results, red - building, blue - road, green - tree.)

Table 1 and Table 2 show two confusion matrices obtained by applying standard CRF model and MSHCRF model to the whole test set respectively. Accuracy values in the table are computed as the percentage of image pixels assigned to the correct class label, ignoring pixels labelled as void in the ground truth. Compared to the confusion matrix showing standard CRF model in Table 1, our MSHCRF model performs significantly better on building and road classes, and slightly better on tree classes. For the similarity in shape and texture between building and road classes in airborne remote sensing images, it is difficult to effectively distinguish them; while the difference in elevation of those classes in LiDAR data can be easily used for classification.

Pr \ Tr	building	road	tree
building	63.7	19.2	17.1
road	22.4	67.0	10.6
tree	11.3	15.2	73.5

Table 1. Pixelwise accuracy of image classification using standard CRF model. The confusion matrix shows classification accuracy for each class (rows) and is row-normalized to sum to 100%. Row labels indicate the true class (Tr), and column labels indicate the predicted class (Pr).

Pr \ Tr	building	road	tree
building	70.1	15.8	14.1
road	14.4	77.3	8.3
tree	12.3	13.8	73.9

Table 2. Pixelwise accuracy of image classification using MSHCRF model. The confusion matrix shows classification accuracy for each class (rows) and is row-normalized to sum to 100%. Row labels indicate the true class (Tr), and column labels indicate the predicted class (Pr).

In our MSHCRF method, we fuse the linear features, such as LF and ASF, and the planar features, such as TF and MSF, to ensure the accuracy of the image classification. In order to verify the necessities of the two kinds of features, we carried out three sets of experiments when retaining both kinds of features, or removing each kind, and Table 3 lists the performance comparison under different conditions. The results show that there are positive effects on the performance of image classification for both kinds of features, in which the linear features is less helpful to the increase of performance because they are difficult to accurately obtain in LiDAR points cloud data for the sparseness and discreteness.

Feature Type	Accuracy(%)
Use both linear and planar features	73.6
Remove the linear features	71.9
Remove the planar features	69.4

Table 3. Drop in overall performance caused by removing each kind of feature in the hierarchical pairwise potentials of MSHCRF model.

5. CONCLUSIONS

In conclusion, this paper presents a novel multi-source hierarchical conditional random field model for feature fusion of remote sensing images and LiDAR data. To exploit the features, category compatibility of images and the category consistency of multi-source data based on the regions selected with typical features, the MSHCRF model is built to classify images into regions of building, road and trees. We have evaluated our approach on airborne data, and the results demonstrate the precision and robustness of the proposed method. For the future work, we are interested in extracting more specific features and corresponding information from the multi-source data to improve the performance of classification.

REFERENCE

Boykov, Y. and Jolly, M.P., 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. *IEEE Conference on Computer Vision and Pattern Recognition*.

Deng, F., Li, S. and Su, G., 2012. Classification of Remote Sensing Optical and LiDAR Data Using Extended Attribute Profiles. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6.

Ge, M., Sun, J., Gao, J. and Wang, Q., 2012. Multiresolution Contrast Modulation Method Applied to Lidar 4-D Image Fusion. *International Conference on Optoelectronics and Microelectronics (ICOM)*, pp. 224-227.

Georgescu, B., Shimshoni, I. and Meer, P., 2003. Mean shift based clustering in high dimensions: A texture classification example. *Proceedings of the International Conference on Computer Vision*，pp. 456-463.

Grompone, R., Jakubowics, J., Morel, J. and Randall, G., 2010. Classification of Remote Sensing Optical and LiDAR Data Using Extended Attribute Profiles. *IEEE Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4), pp. 722-732.

Huang, X., Zhang, L. and Gong, W., 2011. Information fusion of aerial images and LIDAR data in urban areas: vector-stacking, re-classification and post-processing approaches. *International Journal of Remote Sensing*, 32(1), pp. 69-84.

Kadir, T. and Brady, M., 2001. Saliency, Scale and Image Description. *International Journal of Computer Vision*, 45(2), pp. 83-105.

Parmehr, E.G., Zhang, C. and Fraser, C.S., 2012. Automatic Registration of Multi-Source Data Using Mutual Information. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-7.

Shen, W., Zhang, J. and Yuan, F., 2011. A new algorithm of building boundary extraction based on LIDAR data. *19th International Conference on Geoinformatics*.

Shotton, J., Winn, J., Rother, C. and Criminisi, A., 2009. TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. *International Journal of Computer Vision*, pp. 2-23.

Yang, M.Y. and Förstner, W., 2011. A Hierarchical Conditional Random Field Model for Labeling and Classifying Images of Man-made Scenes. *ICCV Workshop on Computer Vision for Remote Sensing of the Environment*, pp. 196-203.