SPARSE CODING-BASED METHOD COMPARISON FOR LAND-USE CLASSIFICATION

Dewa Made Sri Arsa[1], Grafika Jati[1], Yulistiyan Wardhana[1] and M. H. Hilman[2]

[1]Faculty of Computer Science, Universitas Indonesia, Kampus UI, Depok, 16424, Indonesia
[2]Melbourne School of Engineering, University of Melbourne, Building 173, Melbourne, VIC 3010, Australia

E-mail: dewa.made51@ui.ac.id, hilmanm@student.unimelb.edu.au

Abstract

Land-use classification utilize high-resolution remote sensing image. The image is utilized for improving the classification problem. Nonetheless, in other side, the problem becomes more challenging cause the image is too complex. We have to represent the image appropriately. On of the common method to deal with it is Bag of Visual Word (BOVW). The method needs a coding process to get the final data interpretation. There are many methods to do coding such as Hard Quantization Coding (HQ), Sparse Coding (SC), and Locality-constrained Linear Coding (LCC). However, that coding methods use a different assumption. Therefore, we have to compare the result of each coding method. The coding method affects classification accuracy. The best coding method will produce the better classification result. Dataset UC Merced consisted 21 classes is used in this research. The experiment result shows that LCC got better performance / accuracy than SC and HQ. LCC method got 86.48 % accuracy. Furthermore, LCC also got the best performance on various number of training data for each class.

Keywords: *Land-use classification, high-resolution remote sensing image, Bag of Visual Word (BOVW), Sparse Coding (SC), Hard Quantization Coding (HQ)*

Abstrak

Klasifikasi penggunaan lahan memanfaatkan gambar penginderaan jauh beresolusi tinggi. Citra digunakan untuk memperbaiki masalah klasifikasi. Meski begitu, di sisi lain, masalahnya menjadi lebih menantang karena gambarnya terlalu rumit. Kita harus mewakili gambar dengan tepat. Pada metode yang umum untuk mengatasinya adalah Bag of Visual Word (BOVW). Metode ini membutuhkan proses pengkodean untuk mendapatkan interpretasi data akhir. Ada banyak metode untuk melakukan pengkodean seperti Hard Quantization Coding (HQ), Sparse Coding (SC), dan Locality-constrained Linear Coding (LCC). Namun, metode pengkodean itu menggunakan asumsi yang berbeda. Oleh karena itu, kita harus membandingkan hasil setiap metode pengkodean. Metode pengkodean mempengaruhi akurasi klasifikasi. Metode pengkodean terbaik akan menghasilkan hasil klasifikasi yang lebih baik. Dataset UC Merced terdiri dari 21 kelas yang digunakan dalam penelitian ini. Hasil percobaan menunjukkan bahwa LCC memiliki kinerja / akurasi yang lebih baik daripada SC dan HQ. Metode LCC mendapat akurasi 86,48%. Selanjutnya, LCC juga mendapat performa terbaik pada berbagai jumlah data pelatihan untuk masing-masing kelas.

Kata Kunci: *Klasifikasi penggunaan lahan, citra penginderaan jauh beresolusi tinggi, Bag of Visual Word (BOVW), Sparse Coding (SC), Hard Quantization Coding (HQ)*

1. Introduction

Remote-sensing technique has been used as an effective tool to monitor Land-use and land-cover classification. Moreover, remote sensing technique is used to observe dynamic changing of a land [1-3]. Nowadays, single object classification and land classification research are progressive due to the better quality of remote sensing image [4-7].

Land-use-based classification uses image from remote sensing. The image is processed to extract information of land-use. On remote sensing, representation and efficient identification are still open problem and challenging. A lot of previous research used analytical approach, which focused on pixel- or object based classification. It extracted spectral, texture, and geometrical attributes [8-12]. Nevertheless, the attribute is only used in a certain environment so it just produce less data representation.

The recent years, Bag of Visual Words (BOVW) model is implemented to solve Land-use classification problems. It uses remote sensing image data [13]-[15]. Research [13] uses unsuper-vised-

feature-learning approach with Sparse Coding variant that is called Orthogonal Matching Pursuit (OMP-k). Research [14] uses combination of several features. The features was learned using clustering technique. The features are represented in histogram with linear weighting. Research [15] utilize derived method from Sparse Coding, Hard Assignment Vector Quantization. Moreover, research [16] employ Convolutional Neural Network (CNN)-based method named Gradient Boosting Random Convolutional Network (BGRCN). That method use Ensemble CNN which has high complexity of single CNN. Thus, the learning phase takes more time.

Coding as a learning feature and coding has many variations. The variation namely Hard Quantization (HQ), Soft Quantization (SQ), Sparse Coding (SC), Local Coordinate Coding (LCC) , Locality Constrained Linear Coding (LLC), Laplacian Sparse Coding (LSC), Over-complete Sparse Coding (OSC), Saliency Coding (SaC), Super-vector Coding (SV), and Improved Fisher Kernel (IFK) [17]. Each method has different complexity. Bag of Visual Words utilize the coding method to get the data representation.

Land-use classification research usually uses free dataset from UC Merced. The dataset has high degree of difficulty. The dataset has 21 classes of Land-use. This research will compare the performance of several coding methods especially SC, LCC, and HQ for Land-use classification.

The rest of the paper is organize as follows. In the section II, we present method. The section III, result, and analysis are presented. Moreover, we concluded this research in section IV. The last section is the references.

Literature Review

In this part will explain about SC method, HQ, and LCC.

Sparse Coding (SC)
SC method is a method develop from VQ method. SC is a L1-norm regularization for getting a small value that is not 0. Equation(1) shows the sparse coding method.

$$\min_{u,v} \sum_{m=1}^{M} \| x_m - u_m V \|^2 + \lambda | u_m | \qquad (1)$$

That depends on $\| v_k \| \leq 1$, $\forall k - 1, 2, ..., K$. X is a SIFT descriptor and V is a codebook from K clustering.

Locality-constrained Linear Coding (LLC)
The LLC method is initiated by fixing the LCC method, which has a weakness to high computational complexity. This method implements a locality. Therefore, it is important. As a result, The LLC's encoding formula becomes [20] showed in equation(2).

$$\min_{c} \| x_i - CR_i^T \|_2^2 + \lambda \| d_i \Theta R_i \|_2^2 \qquad (2)$$

With
$$d_i = \exp\left[-\beta\left(\| x_i - c_1 \|_2^2 ... \| x_i - c_M \|_2^2\right)\right]$$
which is $\|R_i\|_1 = 1$.

Hard Quantization (HQ)
HQ method presents any local feature with a nearest visual word but only gives good performance when use many vocabularies[21].

$$R_i = \arg\min_{R_i} \| x_i - CR_i^T \|_2^2 \qquad (3)$$

With $Card(R_i) = 1$, $\| R_i \|_2 = 1$, and $R_i > 0$

Besides using different coding techniques, the complexity of each method is also different. The complexity is shown in Table 1. HQ has the highest complexity and SC has the lowest complexity.

TABLE 1
THE COMPLEXITY OF HQ, SC, AND LLC
METHODS

Methods	Complexity
HQ	$O(M)$
SC	$O(M^2)$
LLC	$O(M+K^2)$

2. Methods

In this section, we will describe about the dataset and the method we used on the experiment. We are also present the experiment results and the analysis.

Dataset

To conduct the experiment, we chose to use UC Merced dataset. This dataset is a free data which can be downloaded in http://vision.ucmerced.edu/datasets/landuse.html. It needs to know that this dataset has 21 classes. Figure 2 shows the example of each classes in UC Merced dataset. Each
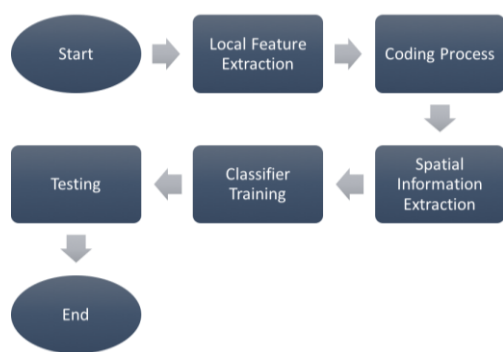
Figure 1. Research method

classes have 100 images, so there are 2100 images for the experiment.

Research Method

Our research method can be seen in Figure 1. There are 5 main process need to be done. The first process is local feature extraction. Then, the coding process is conducted to get the sparse representation of the local feature. After that, the spatial information is extracted from the data based on sparse features. Moreover, the process is continued to the fourth process, classifier training. The final process is testing the performance of our model. Each step will be described below.

Local Feature Extraction
Local feature, extracted from the raw data, is on the image patches form. To extract the local feature, we used Scaled Invariant Feature Transform (SIFT) method. We use dense SIFT to get all information from the data. We set some parameters to be fixed. They are patch size, descriptor degree, and grid spacing. The patch size is set into 16 x 16 px; descriptor has 8 degrees, and grid spacing is set by 8 px. The codebook is set by 1024. Output of this process is descriptor of each patches from the image. Base on this setting, first we extract all of the patches from image. Then, each patch is processed by compute the gradient magnitude.

Coding Process
In this research, we compare the performance of Sparse Coding, Locality constrained Linear Coding, and Hard Quantization method on the coding quality for feature representation in classification task, especially land use classification. The input is the descriptor result from the local feature extraction process. Each local feature will be mapped into sparse representation and locality. The sparse representation means approaching some values close to 0 so that only a few features

are active, whereas locality will provide the feature representation in linear form. This locality makes the final features linearly separated.

Spatial Information Extraction
The result of coding process is a code of local feature for each patch. This result is lacked of spatial information. To address this problem, we used Spatial Pyramid Matching (SPM) method [18]. We divided the image into 3 types of region, 1x1, 2x2, and 4x4. In the 1x1 region, spatial information is extracted on hole image. In the 2x2 region, image will be divided into 4 regions, and 16 regions for 4x4 region type. The function of this division is to eliminate redundant coding features. The input to extract spatial information from the data is the result of the coding process. Then, the result of this partition will be made into one array 1 and the data is ready to be trained using a classifier.

Classifier Training
The classifier used to classify data is the Support Vector Machine (SVM) classifier. Research [13-15] also uses this method as a classification method. In addition, this method is chosen because it is able to maximize margin in the formation of decision boundary.

3. **Results and Analysis**

On the experiment, we measured the accuracy of the classifier. Then, we inspected the influence of data training number.

Classification Accuracy

In here, we divided the training and testing data with ratio 4:1 for each classes. The result can be seen in Figure 3. From this result, LLC performed better than SC and HQ. Because of we used linear classifier, this result proves that LLC has better performance to mapped the local feature into linear space.

The Effect of Amount of Train Data on Accuracy

To know the ability to represent the features in each coding method, the researcher conducted an experiment using different amounts of trainer data. The amount of training data used is 10, 20, 40, and 60. Figure 3 shows the graph of the resulting accuracy. HQ is not involved in comparisons due to HQ dependence on large vocabularies. From Figure 4 it can also be seen that using LLC as a feature encoding provides better accuracy than using the SC method.
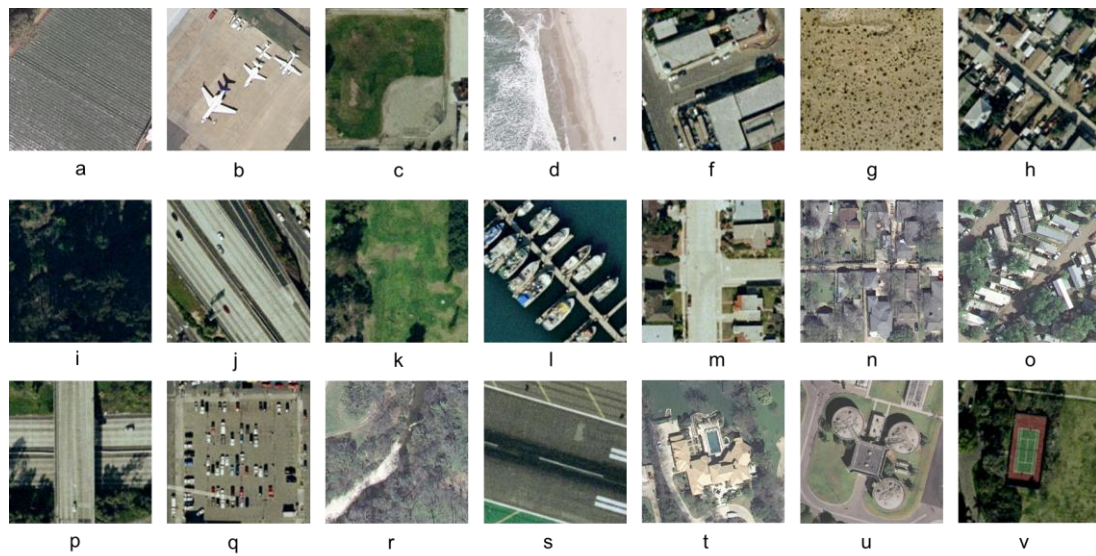
Figure 2. Example of each classes from UC Merced dataset. (a-v) agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile homepark, overpass, parkinglot, river, runway, sparse residential, storage tanks, tennis court.
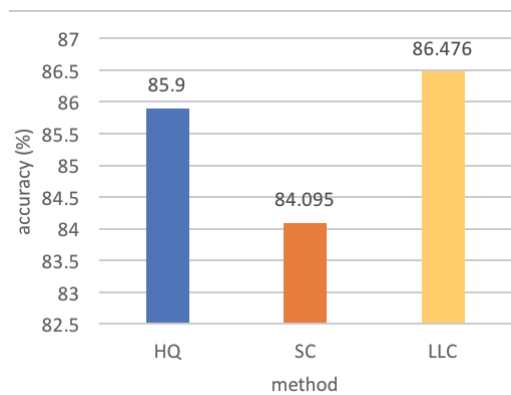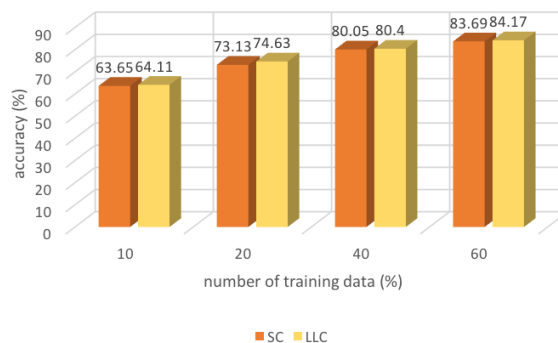

Figure 3. Accuracy of SC, LLC, and HQ

Figure 4. The accuracy of SC and LLC with variation of data training number for each class.

Analysis

From the measurement accuracy of the three methods, it can be seen that LLC has a better ability than SC or HQ. This proves that the

locality that is carried by LLC is important so that it can represent better data. When it comes to land use classification, it relates to the amount of data that can be used as training data. LLC also shows its ability better than SC using little data. However when compared to [16], the accuracy of

LLC is lower. But high accuracy is followed by high complexity in model development.

4. Conclusion

This study has conducted a comparison between HQ, SC, and LLC methods. The measurement results show that LLC has better performance compared to HQ and SC. The number of training data used for the training also determines the accuracy. The more the number of train data, the more improved model recognition capabilities. The highest accuracy was obtained by LLC method of 86.476% for UC Merced dataset.

From the results of this study, it can be done further research which do boosting the method of coding to improve recognition performance. It can also inspect the possibility of other factors besides sparsity and locality that are important in the coding process.

References

[1] R. K. Jaiswal, R. Saxena, and S. Mukherjee, "Application of remote sensing technology for land use/land cover change analysis," *J. Indian Soc. Remote Sens.*, vol. 27, no. 2, pp. 123–128, Jun. 1999.

[2] J. Rogan and D. M. Chen, "Remote sensing technology for mapping and monitoring land-cover and land-use change," *Prog. Plann.*, vol. 61, no. 4, pp. 301–325, May 2004.

[3] Q. H. Weng, J. X. Zhang, P. Gamba, and G. Xian, "Foreword to the issue on remote sensing of regional land use and land cover," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 2, pp. 50–53, Jun. 2009.

[4] F. Palsson, J. R. Sveinsson, J. A. Benediktsson, and H. Aanaes, "Classification of pansharpened urban satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 281–297, Feb. 2012.

[5] A. A. Ursani, K. Kpalma, C. C. D. Lelong, and J. Ronsin, "Fusion of textural and spectral information for tree crop and other agricultural cover mapping with very-high resolution satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 225–235, Feb. 2012.

[6] J. A. dos Santos, P. H. Gosselin, S. Philipp-Foliguet, R. D. Torres, and A. X. Falcao, "Interactive multiscale classification of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 2020–2034, Aug. 2013.

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

[8] M. Pesaresi and A. Gerhardinger, "Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 16–26, Mar. 2011.

[9] I. A. Rizvi and B. K. Mohan, "Object-based image analysis of highresolution satellite images using modified cloud basis function neural network and probabilistic relaxation labeling process," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4815–4820, Dec. 2011.

[10] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C.-W. Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2803–2813, Oct. 2008.

[1]] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354–2363, Oct. 2003.

[12] P. Gamba, F. Dell'Acqua, G. Lisini, and G. Trianni, "Improved VHR urban area mapping exploiting object boundaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2676–2682, Aug. 2007.

[13] A. M. Cheriyadat, "Unsupervised Feature Learning for Aerial Scene Classification," IEEE Trans. Geosci. Remote Sens., vol 52, no. 1, Jan. 2014.

[14] Li-Jun Zhao, P. Tang, and Lian-Zhi Huo, "Land-Use Scene Classification Using a Concentric Circle-Structured Multiscale Bag-of-Visual-Words Model," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 12, Dec. 2014.

[15] S. Chen and YingLi Tian, "Pyramid of Spatial Relatons for Scene-Level Land use

Classification," IEEE Trans. Geosci. Remote Sens., vol 53, no.4, April 2015.

[16] F. Zhang, Bo Du, and L. Zhang,"Scene Classification via a Gradient Boosting Random Convolutional Network Framework," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 3, March 2016.

[17] C. Wang and K. Huang, "How to use Bag of Words Model Better for Image Classification," Image and Vision COmputing, vol. 38, pp. 65-74, 2015

[18] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, Computer Vision and, Pattern Recognition 2006, pp. 2169–2178.

[19] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, Computer Vision and Pattern Recognition 2009, pp. 1794–1801

[20] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, Computer Vision and, Pattern Recognition 2010,pp. 3360–3367.

[21] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, European Conference on Computer Vision International Workshop on Statistical Learning in Computer Vision, 2004.