IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10283228

ACHIEVING 10M CHINA LAND COVER MAPPING WITHIN THREE MINUTES USING A NEW SUNWAY SUPERCOMPUTER

Juepeng Zheng[†,¶], *Yi Zhao*[†,¶], *Jinxiao Zhang*[†,¶], *Wenzhao Wu*[§], *Shuai Yuan*[†,¶], *Haohuan Fu*[†,¶,§]

[†]Ministry of Education Key Laboratory for Earth System Modeling,
and Department of Earth System Science, Tsinghua University, Beijing 100084, China
[¶]Tsinghua University (Department of Earth System Science)- Xi'an Institute of
Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Beijing, China
[§]National Supercomputing Center in Wuxi, Wuxi 214072, China

ABSTRACT

Land Cover Mapping (LCM) is an important task to detect and understand the change of the earth surface. However, most LCM methods adopt supervised classifiers, and suffer from a lack of labels at a large scale. In this paper, we propose Fast-LCM, a scalable and weakly-supervised LCM method on a new Sunway supercomputer to achieve large-scale land cover mapping, requiring no manual annotations. Fast-LCM includes two major parts: (1) a distance-guided k-means module that combines textural, spectral, and temporal features, and (2) an automatic voting-based merging strategy to give each cluster a real meaning of classification system. Through careful parallelization, our Fast-LCM method scales to over 38 million cores, and provides a sustained performance for the task of China LCM. We produce a 10m resolution land cover map of China within only 3 minutes, including 1.2 minutes for IO and only 55 seconds to finish the computation. Fast-LCM achieves an accuracy of 68.85% (25-class), with 3.64% to 7.05% higher than best existing products.

***Index Terms*—** land cover mapping, weakly-supervised, high performance computing, new Sunway supercomputer

1. INTRODUCTION

For the complex surface system of Earth, the rich set of observation satellites serve as a perfect mirror of historical changes. Along with satellite resources, these datasets provide a huge volume of images at different spatial, temporal, and spectral resolutions, to support scientists' efforts on understanding key scientific problems [1]. Of all these different applications to analyze the satellite images, land cover mapping, which categorizes the pixels of the remote sensing images into major land cover types is a basic and important step.

Starting from the 1 km resolution global land cover map [2], the developments in the recent two decades have already promoted the land cover map products to a spatial resolution of 10 m [3] or even 1 m [4]. However, the increased level of complexity in the land cover mapping problem makes it difficult for the above mentioned supervised methods to scale to a large area. To handle the above challenges, we propose Fast-LCM, a scalable and weakly-supervised Land Cover Mapping method on the New Sunway supercomputer.

Correspondence: haohuan@tsinghua.edu.cn (Haohuan Fu)

2. METHODOLOGY

2.1. The Overall Workflow of Fast-LCM

The overall workflow of our Fast-LCM method is shown in Fig. 1. Fast-LCM is a two-stage scheme, consisting of a distance-guided k-means clustering module (Fig. 1(d)) and an automatic voting-based merging strategy (Fig. 1(e)).

Firstly, we present the organization of the entire data before the method details of Fast-LCM. As shown in Fig. 1(a), we decompose of the entire region of China into 23 scenes. Each scene consists of 8×8 tiles and each tile is an image with $13{,}568 \times 13{,}568$ pixels. So the entire China contains $23 \times 8 \times 8 \times 13{,}568 \times 13{,}568 = 2.7 \times 10^{11}$ pixels. To achieve the LCM with huge data amount in a real time, we conduct experiments on a new Sunway Supercomputer.

2.2. A distance-guided k-means module

Here, we take a patch-based design of our k-means approach. As shown in Fig. 1(b), for each pixel, we process an array of the six bands we have selected and another six spectral indexes, including NDVI, NDWI, NDBI, NDMI, MBI and MRBVI. Considering the same scene in satellite images may appear variants in different photographing time, we combine two matrices to cover the two different seasons (*i.e.*, growing season and non-growing season) [5]. By adopting such a patch-based design (a detailed arrangement of the data items can be seen in Fig. 1(c)), we are able to integrate textural, spectral, and temporal patterns into a unified feature vector, with 600 dims of feature for each pixel.

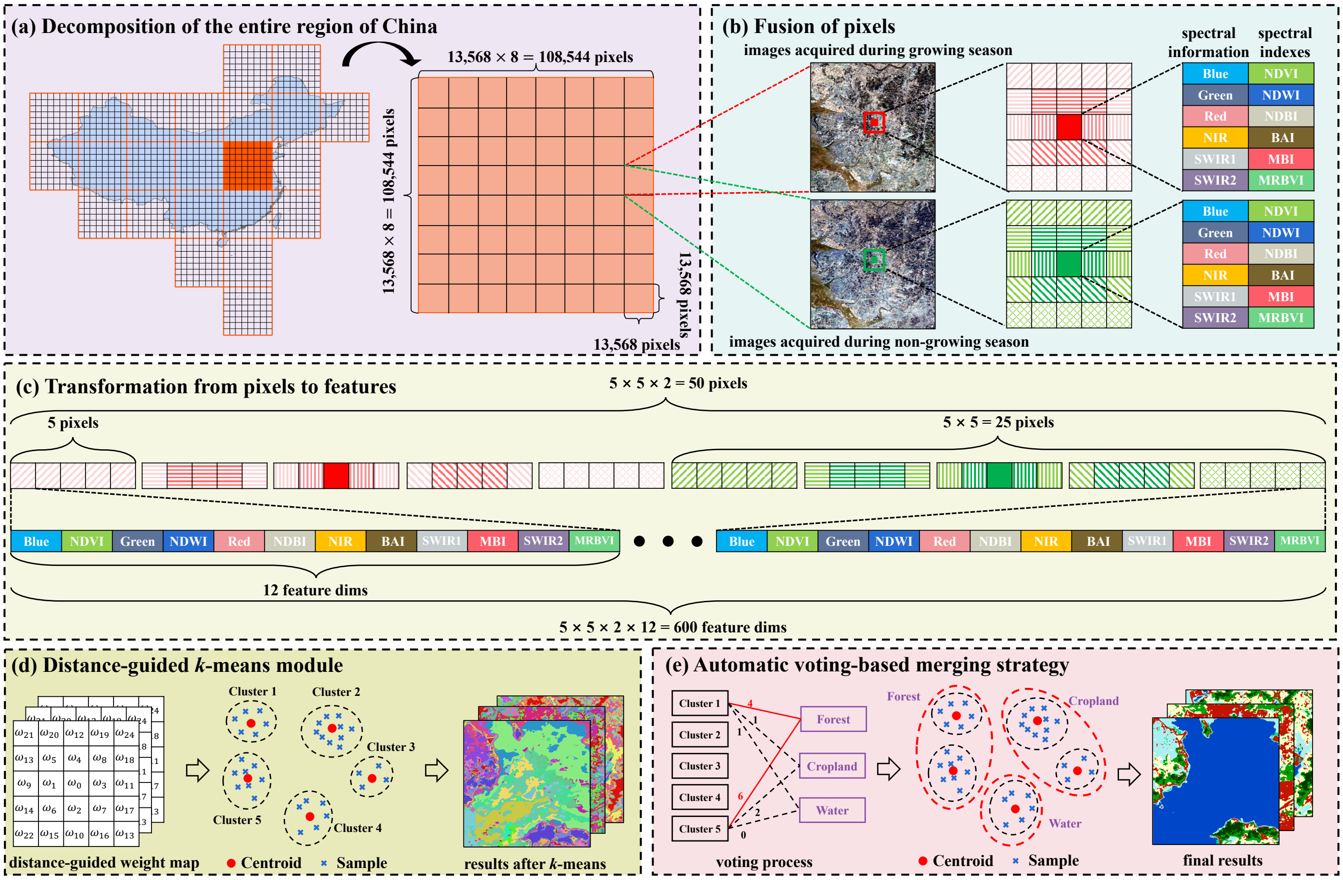

Fig. 1. Workflow of our proposed Fast-LCM. (a) decomposition of the entire region of China into 23 scenes with 2.7×10^{11} pixels in total; (b) Fusion of pixels with 6 spectral information and 6 spectral indexes; (c) Transformation from pixels to features with 600 feature dims; (d) distance-guided k-means that combines textural, spectral, and temporal features; (e) automatic voting-based merging strategy to give each cluster a real meaning of classification system.

However, although the patch-based design in k-means makes the land cover mapping results more robust, assigning equal weight for the neighborhood in the patch is unreasonable. Here, we design a distance-guided k-means that focuses on the central pixel and the weights will decay exponentially according to the distance to the central pixel, which can alleviate the noise information from the patch.

2.3. An automatic voting-based merging strategy

It is apparently not an easy task to guide the k-means generated classes to converge to the 25 target classes. To resolve this issue, we choose to use a k that is much larger than 25 (300 in our case), and then merge the clustered classes into the 25 target classes with guidance from existing products. Therefore, we need to carefully map between the machine-generated classes and the expert-defined classes.

2.4. Parallelization

Fast-LCM is implemented with elaborate optimization and parallelization to fully utilize the great computing power of modern heterogeneous architectures such as the new Sunway supercomputer. A series of performance optimization strategies are applied in Fast-LCM to squeeze the computing power including reasonable task and data allocation, mixed-precision computation, efficient loop reorganization, and better memory access and usage pattern. Furthermore, based on these optimization, Fast-LCM can efficiently scale to up to 589,824 processes on the new Sunway supercomputer via MPI, completing land cover mapping of China within 3 minutes.

3. EXPERIMENTAL RESULTS

3.1. Scaling Results

Fig. 2 shows the strong and weak scaling results for Fast-LCM on the new Sunway supercomputer. Fast-LCM effi-

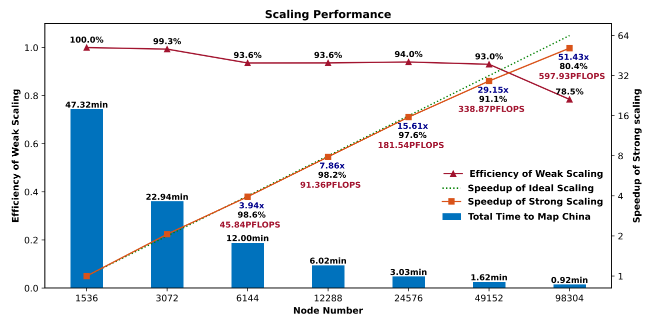

Fig. 2. Strong and weak scaling performance

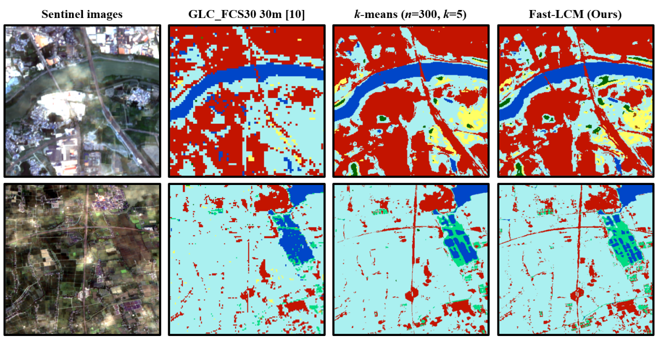

Fig. 3. We display more detailed land cover mapping results among different products and methods with 25-class.

ciently scales to up to 589,824 processes running on 98,304 nodes in the experiments of both strong and weak scaling test. Each process is assigned to a Core Group. The Fast-LCM demonstrates highly-efficient parallel performance and reduces the time needed to map the entire region of China from 47.3 minutes to less than one minute to finish the computation, providing a sustained performance of 597.9 Pflops.

3.2. Accuracy of Fast-LCM

To evaluate our results, we manually annotate 7,341 sites as the validation dataset. As shown in Table 1, Fast-LCM achieves an accuracy of 68.85% (25-class), with 3.64% to 7.05% higher than other products for 25-class. As far as we know, this work produces the first 10-m fined-grained land cover map for the entire region of China with 25 classes in the year of 2020. Here, we display more detailed land cover mapping results in Fig. 3. Fast-LCM is more consistent with the ground truth.

4. CONCLUSION

Here, we propose Fast-LCM, a scalable and weakly-supervised land cover mapping method on a new Sunway supercomputer. Fast-LCM consists of a distance-guided k-means module that combines textural, spectral, and temporal features, and the

Table 1. The accuracy (%) of land cover mapping

Methods	Accuracy
ESA CCI LC [6]	61.80
GLC_FCS30 [7]	65.21
(Fast-LCM)	**68.85**

other one is an automatic voting-based merging strategy to give each cluster a real meaning of classification system. With a careful design of both parallelization and optimization strategies, we manage to scale our method to 98,304 nodes for k-means (597.9 Pflops) on the new Sunway supercomputer. Overall, we can finish the land cover mapping of China within 3 minutes, with 1.2 minutes for IO, and only 55 seconds to finish the computation. The Fast-LCM achieves an accuracy of 68.85%. This is the first efforts to achieve a 10m resolution land cover map with 25-class.

5. ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Plan of China (Grant No. 2020YFB0204800), National Natural Science Foundation of China (Grant No. T2125006) and Jiangsu Innovation Capacity Building Program (Project No. BM2022028).

6. REFERENCES

[1] Juepeng Zheng, Shuai Yuan, Wenzhao Wu, Weijia Li, Le Yu, et al., "Surveying coconut trees using high-resolution satellite imagery in remote atolls of the pacific ocean," *Remote Sensing of Environment*, vol. 287, pp. 113485, 2023.

[2] Matthew C Hansen, Ruth S DeFries, et al., "Global land cover classification at 1 km spatial resolution using a classification tree approach," *International journal of remote sensing*, vol. 21, no. 6-7, pp. 1331–1364, 2000.

[3] Peng Gong, Han Liu, Meinan Zhang, Congong Li, Jie Wang, Huabing Huang, Nicholas Clinton, et al., "Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017," *Sci. Bull*, vol. 64, pp. 370–373, 2019.

[4] Caleb Robinson, Le Hou, Kolya Malkin, Rachel Soobitsky, Jacob Czawlytko, Bistra Dilkina, and Nebojsa Jojic, "Large scale high-resolution land cover mapping with multi-resolution data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12726–12735.

[5] Juepeng Zheng, Yi Zhao, Wenzhao Wu, Mengxuan Chen, Weijia Li, and Haohuan Fu, "Partial domain adaptation for scene classification from remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[6] Esa, "Esa: Land cover cci product user guide version 2.0," 2017.

[7] Xiao Zhang, Liangyun Liu, et al., "Glc_fcs30: Global land-cover product with fine classification system at 30 m using time-series landsat imagery," *Earth System Science Data*, vol. 13, no. 6, pp. 2753–2776, 2021.