MULTIMODAL, MULTITEMPORAL, AND MULTISOURCE GLOBAL DATA FUSION FOR LOCAL CLIMATE ZONES CLASSIFICATION BASED ON ENSEMBLE LEARNING

Naoto Yokoya[,†,‡], Pedram Ghamisi[†,‡], and Junshi Xia[*]*

[*]Department of Advanced Interdisciplinary Studies, the University of Tokyo, Japan
[†]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany
[‡]Signal Processing in Earth Observation (SiPEO), Technische Universität München (TUM), Germany

ABSTRACT

This paper presents a new methodology for classification of local climate zones based on ensemble learning techniques. Landsat-8 data and open street map data are used to extract spectral-spatial features, including spectral reflectance, spectral indexes, and morphological profiles fed to subsequent classification methods as inputs. Canonical correlation forests and rotation forests are used for the classification step. The final classification map is generated by majority voting on different classification maps obtained by the two classifiers using multiple training subsets. The proposed method achieved an overall accuracy of 74.94% and a kappa coefficient of 0.71 in the 2017 IEEE GRSS Data Fusion Contest.

Index Terms— Local climate zones (LCZs), canonical correlation forests, rotation forests, morphological profiles.

1. INTRODUCTION

In recent years, there has been a growing interest in the mapping of local climate zones (LCZs) for urban temperature studies [1]. LCZs comprises 17 classes based on properties of 3D surface structure (e.g., height and density of buildings and trees) and surface cover (e.g., previous or impervious). The classification of LCZs is a challenging task due to a large intra-class variability of spectral signatures caused by the regional variations of vegetation and artificial materials.

The LCZ classification scheme has been successfully developed via the World Urban Database and Access Portal Tools (WUDAPT[1]) to collect information regarding the form and function of cities worldwide [2]. The contributors of WUDAPT create training data using high-resolution images obtained by Google Earth. The LCZ classification is performed on Landsat data using random forests [3] in a System developed for Automated Geoscientific Analyses (SAGA GIS). Although the scheme has been well established, it still

requires an enormous amount of human efforts, particularly for the specific step of creating training data.

To overcome this limitation, we promote innovative classification algorithms, which have high generalization ability and transferability. Therefore, the Image Analysis and Data Fusion Technical Committee (IADFTC) of the IEEE Geoscience and Remote Sensing Society (GRSS) organized the contest of the LCZ classification in 2017, aiming at accelerating progress on data fusion methodologies for multitemporal, multisource, and multimodal remote sensing data.

We present here a new multimodal data fusion methodology for the LCZ classification based on ensemble learning techniques. Canonical correlation forests (CCFs) [4] and rotation forests (RoFs) [5] are applied to spectral-spatial features extracted from satellite images and semantic layers. Our results achieved an overall accuracy of 74.94%, which ranked first in the contest among more than 800 submissions.

The remainder of the paper is organized as follows. Section II describes the data sets provided by the fusion committee for the contest. Section III introduces our proposed methodology. Experimental results and discussion are presented in Section IV. Section V concludes the paper.

2. DATA SETS

The data sets comprise multitemporal, multisource, and multimodal data for nine cities: five cities (i.e., Berlin, Rome, Paris, Sao Paulo, and Hong Kong) for training and four cities (i.e., Amsterdam, Chicago, Madrid, and Xi'an) for testing. For each city, the following three kinds of image data have been provided for the contest.

- Landsat data[2]: Eight multispectral bands (i.e., visible, short and long infrared wavelengths) resampled at a ground sampling distance (GSD) of 100 m with multitemporal (i.e., 2-6) acquisitions.

- Sentinel-2 data: Nine multispectral bands (i.e., visible, vegetation red edges and short infrared wavelengths) resampled at a 100 m GSD for a single acquisition.

[1]http://www.wudapt.org/

[2]Courtesy of the U.S. Geological Survey
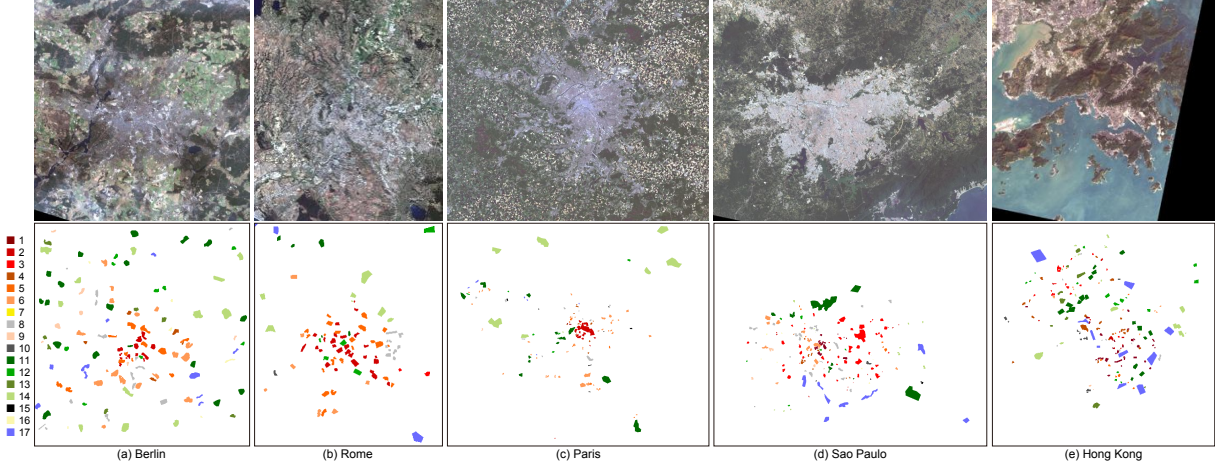
Fig. 1. Color composite images (top) and ground truth (bottom) for five training cities: (a) Berlin, (b) Rome, (c) Paris, (d) Sao Paulo, and (e) Hong Kong.

Table 1. LCZs and numbers of pixels for training and test.

No.	Color	Name	Train	Test
1		Compact high-rise	1642	242
2		Compact mid-rise	6103	4892
3		Compact low-rise	5738	1522
4		Open high-rise	2098	2270
5		Open mid-rise	4759	2255
6		Open low-rise	8891	8265
7		Lightweight low-rise	0	0
8		Large low-rise	4889	11230
9		Sparsely built	1156	1072
10		Heavy industry	449	920
11		Dense trees	17716	3170
12		Scattered trees	2819	4528
13		Bush, scrub	1741	1284
14		Low plants	14457	12994
15		Bare rock or paved	323	1104
16		Bare soil or sand	503	391
17		Water	8561	4454

- Open Street Map (OSM) data[3]: Three layers of "buildings," "land-use," and "water" at a 5 m GSD.

For the training cities, ground-truth samples of the LCZ classes were provided in several areas of each city in the form of raster images at a 100 m GSD. Table 1 summarizes class names and colors of the LCZ classes and the numbers of pixels for training and test. Fig. 1 shows the color composite images created from the Landsat-8 data as well as the ground-truth maps for the training cities.

Among the data mentioned above, only Landsat and OSM data have been used in this work. We used Landsat data but not Sentinel-2 data to exploit long infrared wavelengths. By using the available time-series of the Landsat data, one can take temporal-spectral variability into consideration to effectively train the classifiers. It should be noted that the original Landsat-8 images at a GSD of 30 m were downloaded via Amazon Simple Storage Service (Amazon S3) for all images provided in the contest. All eleven bands were used as input, which is detailed in the next section.

---

[3]Data © OpenStreetMap contributors, available under the Open Database Licence http://www.openstreetmap.org/copyright

3. METHODOLOGY

Fig. 2 illustrates the flowchart of our algorithm developed for the classification of LCZs in the framework of the 2017 IEEE GRSS Data Fusion Contest. The algorithm is mainly composed of four steps: preprocessing, feature extraction, classification, and postprocessing as detailed below. Due to the fact that the number of available data sets and the corresponding training samples provided by the fusion committee was high, a particular emphasis in the proposed framework was dedicated to fast, automatic, yet effective approaches to achieve accurate results in an acceptable CPU processing time.

3.1. Preprocessing

Atmospheric correction was performed on the original Landsat-8 images using ATCOR-2/3 version 9.0.0 with the haze removal option. The Landsat-8 images were upsampled at a GSD of 10 m using bicubic interpolation. The subareas used in the contest have been extracted using sub-pixel precision image matching based on phase correlation at a GSD of 100 m. All OSM images have been normalized between 0 and 1 and spatially downgraded to a GSD of 10 m to reduce the computational cost in the subsequent processing steps.

3.2. Feature Extraction

The Landsat-8 images as well as the OSM images have been investigated to extract input features suitable for the classification of LCZs. A total number of 44 features, which are composed of spectral reflectance, spectral indexes, and spatial features, have been extracted at a GSD of 100 m as described in detail below:

- Mean and standard deviation have been calculated for each patch of 10×10 pixels for all bands of the 10 m-GSD Landsat-8 data (22 features).
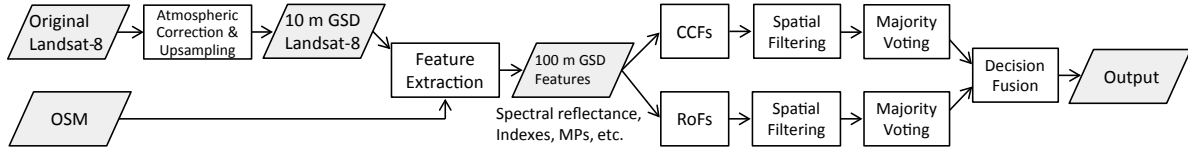
Fig. 2. Flowchart of the classification algorithm.

- Mean and standard deviation have been calculated for each patch of 10×10 pixels for three indexes obtained from the 10 m-GSD Landsat-8 data, namely, the normalized difference vegetation index (NDVI), the normalized difference water index (NDWI), and the bare soil index (BSI) (6 features). The advantage of using these indexes has been already shown for LCZ classification [2].

- Mean has also been calculated for each patch of 10×10 pixels for the 10 m-GSD OSM images of "buildings," "land-use," and "water" (3 features).

- Morphological profiles (MPs) [6] composed of opening and closing by reconstruction have been used to extract spatial information from the 10 m-GSD NDVI and OSM "building" images. To do so, a circular structuring element with threshold values of 1, 2, and 3 have been taken into account (14 features).

3.3. Classification

Two ensemble learning methods, namely CCFs[4] [4] and RoFs [5], were used for the classification step. For rotation of feature axes before constructing decision trees, CCFs adopt canonical correlation analysis between features and labels, whereas RoFs use random splits of features and principal component analysis. The superior performance of these two ensemble classifiers have been already proven in the remote sensing community in terms of classification accuracy, CPU processing time, and generalization capability [7].

The number of trees was set to 20 for both methods with reference to the studies reported in [7]. 15 different training data sets were prepared. The first ten sets were created by splitting the whole training data provided by the fusion committee into ten subsets whereas the other five sets were obtained by randomly extracting the same number of training samples (i.e., 500) for all classes. In this way, one can increase the diversity of the forests, which plays an important role to boost the classification performance of ensemble and multi-classifiers. Although some classes include fewer than 500 training pixels, the total number of training samples for all classes is over 500 since each city has multiple Landsat-8 images. It should be noted that some Landsat-8 images that include clouds were not used for the training step. Both CCF and RoF were built on each training data set, resulting in 30

---

[4]The source code is available at https://bitbucket.org/twgr/ccf

Table 2. OA and kappa for two versions of training data.

Sampling of training data	OA (%)	Kappa
Uniform	70.29	0.66
Uniform & imbalance-corrected	74.94	0.71

different forests in total. Classification was performed on selected cloud-free images for each city using CCFs and RoFs.

3.4. Post-processing

Spatial filtering was applied to all classification maps using a 3×3 median filter to reduce the labeling uncertainty and salt and pepper appearance of labeled pixels. The final classification map was obtained using majority voting on $15 \times N$ classification maps obtained by each ensemble learning method, where N is the number of Landsat-8 images for each city. For the final classification maps, the results of CCFs were used for Amsterdam, Chicago, and Xi'an, whereas those of RoFs were used for Madrid.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 3 shows the LCZ classification maps for the test cities obtained by the proposed algorithm. Our result achieved an overall accuracy (OA) of 74.94% and a kappa coefficient of 0.71. Table 2 shows OA and kappa for two versions of our results. The first one was obtained by using the first ten training data sets with uniform sampling (i.e., different numbers of training samples for different classes). The second one demonstrates our final result obtained by merging the results with the other five training data sets, which include the same number of training samples for all classes (i.e., imbalance-corrected sampling). By integrating the classification results obtained using the additional five subsets of training data, the classification accuracy of minor classes has been increased, leading to the improvement of the overall result.

Table 3 presents the confusion matrix with the producer's accuracy (PA) and the user's accuracy (UA) for the final result to further analyze the performance of our algorithm. The classification accuracies of classes 6, 8, 11, 14, 16, and 17 are relatively high, achieving over 75%. Many of those classes include a high number of training samples as shown in Table 1. For instance, classes 6, 11, 14, and 17 have more than 8000 pixels for training. These results demonstrate that high
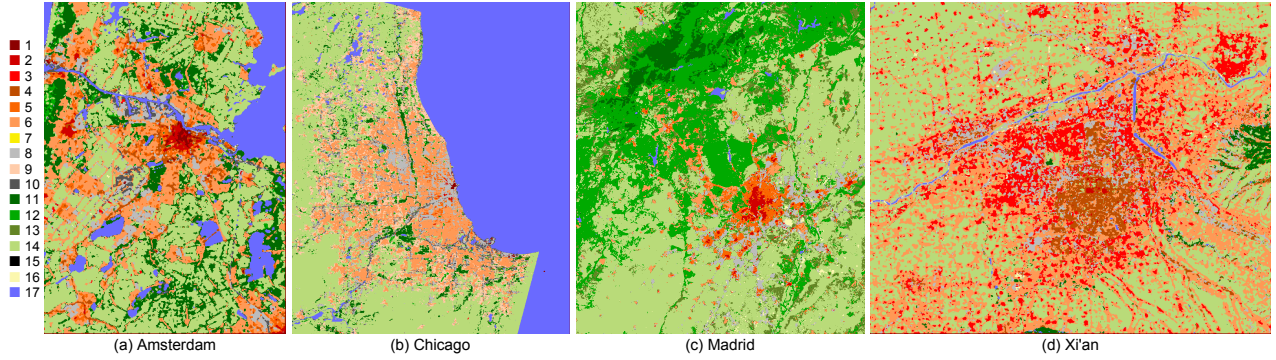
Fig. 3. LCZ maps for four test cities: (a) Amsterdam, (b) Chicago, (c) Madrid, and (d) Xi'an.

Table 3. Confusion matrix.

		1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	UA (%)
								Ground truth										
Prediction	1	167	151	15	1	1	2	20	0	24	0	0	0	0	0	1	0	43.72
	2	2	3033	321	97	166	41	16	0	3	0	0	0	0	0	0	0	82.44
	3	0	181	624	125	6	230	375	0	46	0	0	0	84	67	0	0	35.90
	4	12	350	18	1243	121	52	81	0	8	1	15	0	1	0	0	9	65.04
	5	0	469	202	242	1540	1223	74	0	1	5	114	0	1	11	0	0	39.67
	6	1	343	168	250	252	6432	1004	86	85	65	227	12	330	25	0	16	69.19
	8	59	294	149	290	154	89	9055	0	625	25	66	0	51	429	15	80	79.56
	9	0	0	0	2	0	13	124	18	1	10	10	0	7	13	0	11	8.61
	10	0	0	0	15	1	5	305	1	92	5	21	0	3	107	16	23	15.49
	11	0	0	0	0	0	6	14	28	20	2690	201	11	207	2	0	3	84.54
	12	0	2	2	0	6	50	7	0	0	190	3230	131	454	300	0	6	73.78
	13	0	11	2	0	7	66	3	10	0	2	391	940	57	11	0	0	62.67
	14	1	57	18	2	1	55	96	929	6	177	233	190	11788	29	26	28	86.45
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—
	16	0	0	16	0	0	0	45	0	9	0	0	0	1	78	294	6	65.48
	17	0	1	0	3	0	1	11	0	0	0	20	0	10	32	39	4272	97.33
PA (%)		69.01	62.00	40.65	54.76	68.29	77.82	80.63	1.68	10.00	84.86	71.33	73.21	90.72	0.00	75.19	95.91	

generalization ability and transferability of classifiers can be obtained by feeding sufficient training data to the algorithm.

On the other hand, the classification accuracies of classes 3, 4, 9, 10, and 15 are poor (less than 60%). There are two different reasons: (1) For classes 10 and 15, the number of training samples is relatively limited, which makes it impossible to learn intra-class variability in an effective way. As a result, it is difficult to increase the generalization ability and transferability of the classifiers. (2) Classes 3, 4, and 9 are misclassified into classes whose spectral-spatial features are similar. For instance, the following class pairs are confused: classes 2 and 3, classes 4 and 5, classes 9 and 14. To distinguish these pairs, a digital surface model is a key data source.

5. CONCLUSION

We proposed a new methodology for LCZ classification based on ensemble learning techniques. CCFs and RoFs are applied to classify spectral-spatial features (e.g., spectral reflectance, spectral indexes, and MPs) extracted from the Landsat-8 and OSM data. Our algorithm achieved 74.94% OA in the 2017 IEEE GRSS Data Fusion Contest.

Our future work will be on analysis of feature importance. Further study on investigating the impact of using recent advances in spatial feature extraction (e.g., extinction profiles [8]) will also be an interesting line of research.

7. REFERENCES

[1] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bulletin of the American Meteorological Society*, vol. 93, no. 12, pp. 1879–1900, 2012.

[2] B. Bechtel, P. J. Alexander, J. Böhner, J. Ching, O. Conrad, J. Feddema, G. Mills, L. See, and I. Stewart, "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS Inter. Jour. Geo-Information*, vol. 4, no. 1, pp. 199–219, 2015.

[3] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[4] T. Rainforth and F. Wood, "Canonical correlation forests," *ArXiv Preprint ArXiv:1507.05444.*

[5] J. J. Rodriguez and L. I. Kuncheva, "Rotation forest: A new classifier ensemble method.," *IEEE Trans. Pat. Anal. Mach. Intel.*, vol. 28, no. 10, pp. 1619–1630, 2009.

[6] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geos. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.

[7] J. Xia, N. Yokoya, and A. Iwasaki, "Hyperspectral image classification with canonical correlation forests," *IEEE Trans. Geos. Remote Sens.*, vol. 55, no. 1, pp. 421–431, 2017.

[8] P. Ghamisi, R. Souza, J. A. Beneiktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geos. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.