EFFICIENT BUILDING CATEGORY CLASSIFICATION WITH FAÇADE INFORMATION FROM OBLIQUE AERIAL IMAGES

C. Xiao[1,3], X. Xie[2,3]*, L. Zhang[4], B. Xue[2,3]

[1] Artificial Intelligence and Earth Perception Research Center, School of Automation Engineering, University of Electronic Science and Technology of China, China - channingshaw@hotmail.com
[2] Key Lab of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China -xiexiao@iae.ac.cn
[3] Key Lab for Environmental Computation and Sustainability of Liaoning Province, Shenyang 110016, China
[4] Department of compute science and engineering ,Southern University of Science and Technology - zhangl33@sustech.edu.cn

Commission II, WG II/III

KEY WORDS: Building Category, Façade, Oblique Aerial Images, Remote Sensing, Classification

ABSTRACT:

Building category refereed to categorizing structures based on their usage is useful for urban design and management and can provide indexes of population, resource and environment related problems. Currently, the statistics are mainly collected by manual from street data or roughly extracted from remote sensing data which are either laborious or too coarse. With remote sensing data (e.g. satellite and aerial images), buildings can be automatically identified from the top-view, but the detailed categories of single buildings are not recognized. Façade from oblique-view image can greatly help us to identify the categories of buildings, for example, balcony usually exist in resident buildings. Hence, in this paper, we propose an efficient way to classify building categories with the façade information. Firstly, following the texture mapping procedure, each building's façade textures are cropped from oblique images via a perspective transformation. Then, the average colour, the stander deviation in R, G, B channel, and the rectangle Haar-like features are extracted and feed to a further random forest classifier for their category identifications. In the experiment, we manually selected 262 building façades that can be classified into four functional types as: 1) regular residence ; 2) educational building; 3) office ; 4) condominium. The results shows that, with 30% data as training samples, the classification accuracy can reach 0.6 which is promising in real applications and we believe with more sophisticated feature descriptors and classifiers, e.g., neuronal networks, the accuracy can be much higher.

1. INTRODUCTION

The building category refereed to categorizing structures based on their usage is useful for urban design and management. This kind of information can provide indexes of population, resource and environment-related problems, such as population distribution, power supply, and traffic system design. They are the basis of urban planning, policy-making and disaster management (Kolbe et al., 2005, Tutzauer et al., 2016). Currently, the statistics are mainly collected by manual from street data or roughly extracted from remote sensing data which are either laborious or too coarse. With remote sensing data, some land use classification methods (Rawat, Kumar, 2015, Liu et al., 2016, Zhang et al., 2015a) can automatically identify the airport from residential area or the industrial zone from public facilities, but the detailed categories of individual buildings are not recognized. On the other hand, single buildings can be classified and detected from overview data, such as satellite and aerial images. However, most of these building detection methods use the top-view information such as the appearance of the roof and the high from DSM (digital surface model) which are not enough for individual building category identification. Hence, how to efficiently acquire such information at a large scale (e.g. city) is still a problem.

With the development of multi-camera/head imaging systems, many remote sensing platforms can simultaneously capture the top-view and oblique-views images in different directions. This oblique imagery is widely used for photogrammetric 3D reconstruction, its cartographic mapping products (orthophoto and DSM) are becoming popular for land-cover classification (Zhang et al., 2015b), building detection and modelling (Gruen et al., 2019). Besides building geometric information, the oblique images also contain façade textures that are helpful for building category identification, such as balcony indicating residence while a large part of glass showing an office. A building's 3D geometric representation usually contains semantic information such as building category, architectural style, and historical relevance. The analyses of the study (Tutzauer et al., 2016) reveal clear coherence and dependencies between the correctness of classifications and the model representation types.

Hence, in this study, with the oblique images, we propose a framework to efficiently identify building categories at a large scale based on their façade textures. Firstly, from existing building footprints or LoD2 (level-of-detail 2) building models, the corresponding building boundaries can be registered and matched in orthophoto and DSM derived from the oblique images. Then, based on the geo-referenced coordinates of buildings, the façade textures can be cropped and selected from oblique images, as same as the texture mapping procedures in (Xiao et al., 2020). Finally, from these façade textures, color and texture features are extracted and further fed to a random forest classifier for the building category classification.

* Corresponding author

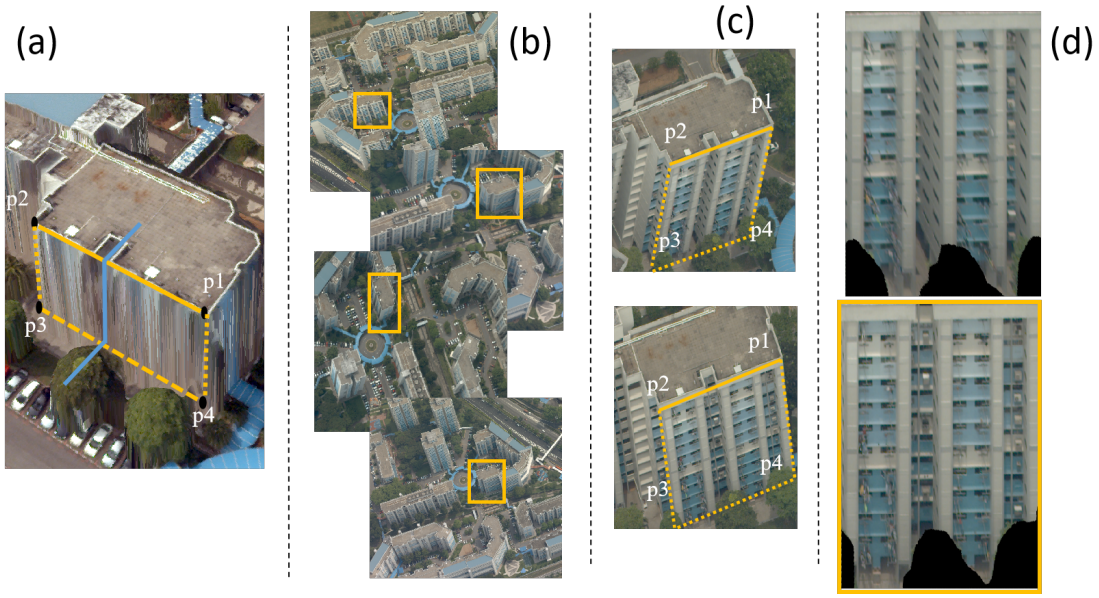

Figure 1. The façade texture extraction from multi-view oblique images. Image (a) shows a buildings' 3D vertical face and the possible projections at multi-view oblique images are showing in image (b). Finally, image (c) and (d) show the original and rectified façade textures while the yellow rectangle in image (d) marks the best texture.

2. RELATED WORK

Usually it is difficult to determine the category of the buildings in complicated urban areas just from the top-view. Hence, many researches are using ground-level images as supplementary data for the land (mainly the man-made facilities) use category classification (Newsam, 2010, Zhu, Newsam, 2015). In (Zhu et al., 2018), building photos, including intern scenes, are collected and labelled to train a deep convolutional neural network (CNN) for building category classification. Then the image with geolocation information can be classified and further to help identify the location's land use type. However, due the geolocation error, this method also only offer patch or area based land use classification. And this method is heavily depended on the availability of photos that has geo-information. Besides ground-level images, some researches are using point of interest information from social network which contains the functional and locational properties (Ty et al., 2016, Deng, Newsam, 2017). Even these methods work well in cities, it is still an approximate estimate.

Single image of the top-view is limited to offer the use information of the buildings, but the temporal sequences of images and other metadata can provide clues for the land/building use classification. Recently, the functional map of the world (fMoW) challenge was launched to ask resolutions to classify facility, building, and land use from satellite imagery. Their baseline method is using Long Short-Term Memory (LSTM) neural network and metadata to identify different type of buildings including hospital, office building, police station, and etc (Christie et al., 2017). Also, some other deep learning neural networks are adopted to deal this kind of problem which including using ensemble convolutional neural networks (Minetto et al., 2019). However, for the building category classification, the accuracy is relatively low, such as the office building only have 0.225 accuracy in the baseline method. Also, most of the classifications are at patch or area scale, not the individual building.

A building's 3D geometric representation usually contains semantic information such as building category, architectural style and historical relevance. To explore the connections between the 3D geometric and building category, (Tutzauer et al., 2016) developed a tool to ask user to classify buildings into six characteristic building categories. And the analyses of the study reveal clear coherences and dependencies between the correctness of classifications and the model representation type.

Unlike all above method, we explore the façade information from oblique images which capture some kind of semantic information of buildings to help us identify the category of buildings at the individual level. The façade can be exacted from oblique images which are cheap and conveniently to be acquired.

3. FAÇADE FEATURE EXTRACTION AND CLASSIFICATION

From existing building footprints or LoD2 (level-of-detail 2) building models, the corresponding building boundaries can be registered and matched in orthophoto and DSM derived from the oblique image. Similar to 3D building façade texture mapping (Frueh et al., 2004), the vertical faces of above-ground objects can be mapped and cropped from oblique images. Our previous work about urban land-cover classification with façade information (Xiao et al., 2020) has provided a pipeline for the texture mapping and cropping from multi-view oblique images. Hence, we directly adopt its procedure as the following processes.

3.1 Façade Texture Mapping

One building usually has more than one façade, in this study, we selected the longest façade as its representation. As illustrated in Figure 1, image (a), the vertical face is defined as a rectangle with four space points $(P1, P2, P3, P4)$. The upper points $(P1, P2)$ are the two ending points of a polygon line with the object height, while the lower points $(P3, P4)$ are at the same positions but with ground height. The georeferenced

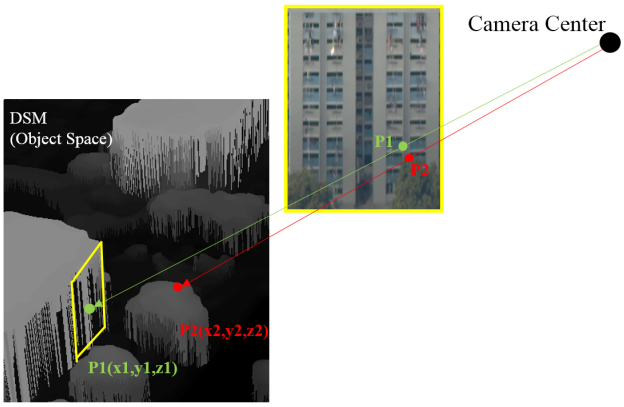

Figure 2. An illustration of the occlusion detection through Z-buffer with the DSM. A texture point (e.g. $p1$), must be close to the façade plane (yellow rectangle) in the object space, otherwise (e.g. $p2$ which is pointing at a tree) it should be an occlusion point.

3D coordinates (X, Y, Z) of the four points in the object space can be acquired from the orthophoto and DSM, thus their corresponding oblique image coordinates can be calculated via a perspective transformation:

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = P_{3\times4} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad (1)$$

where $(u, v, 1)$ is the 2D homogeneous coordinates in the oblique image with s as a scale factor, and $P_{3\times4}$, is a perspective transform matrix which contains the intrinsic and extrinsic camera parameters that are calibrated in the photogrammetric 3D processing. The reader can find more details about the photogrammetry in (Hartley, Zisserman, 2003). As illustrated in Figure 1, image (b) and (c), after this perspective transform, the four points can define a region of the façade in each multi-view oblique image. To get better façades for the later feature extractions, we rectify the textures to the front view through a homography transform that maps the points in one image to the corresponding points in the other image (e.g. mapping $P2$, $P1$, $P3$, $P4$ to the top-left, top-right, bottom-left, bottom right corner of a rectangle image, separately), as shown in Figure 1 (d). The readers can find more details about homography in (Hartley, Zisserman, 2003).

There are in general more than one oblique images can capture a façade of an object. To select the best one, we consider three factors: 1) $V(f)$, the quality of the angle between the normal of the face plane and the camera imaging plane, 2) $N(f)$, the quality of the angle between the face normal and the line through camera and face centers, 3) $O(f)$, the proportion of the observable part. Based on these factors, the best façade is selected by a texture quality measurement:

$$Q(f) = m_1 * V(f) + m_2 * N(f) + m_3 * O(f), \quad (2)$$

where the $Q(f)$ measures the quality of façade f, while the m_1, m_2, m_3 are the weights of different quality factors. In the experiment, m_1, m_2, m_3 are set as 0.25, 0.25, 0.5, respectively, as we found the visibility is more important. While the first two factors can be easily calculated, the visibility is complicated to measure due to that the occlusion often exists in urban areas. Inspired by a Z-buffer based occlusion detection (Rau et al., 2014), we examine the visibility with a distance measurement as illustrated in Figure 2.

For each façade region in the multi-view oblique images, we can simulate emitting rays from its camera center through the façade texture and reach the DSM in the object space. If a pixel is not part of the plane (e.g. due to occlusion), like $P2$ in Figure 2, we will determine that as an invalid pixel for feature extraction. The resulting masked image is shown in image (d) of Figure 1.

3.2 Façade Feature Description and Classification

To capture the façade features, we take segmented images and compute the average color and the standard deviation in R, G, B channels. Considering that the elements (e.g., windows) in the building façades usually have a regular and repetitive layout, we adopt the rectangle Haar-like features (Crow, 1984, Viola et al., 2001) to the façade images, as has been shown to be highly descriptive. The rectangle Haar-like feature is defined as the difference of the sums of the pixel intensities inside different rectangles. For the façade textures, a triple-rectangle pattern Haar-like structure(e.g. black-white-black) is designed and used at the vertical and horizontal direction, separately, at 3 different sizes (total 6 feature vectors). Finally, from pixels to blocks, the color and Haar-like features are combined to describe the façade for each building. The random forest(RF) classifier is widely used for hierarchical feature classifications (Sun et al., 2017). The voting strategy of multiple decision trees and the hierarchical examination of the feature elements make this method have high accuracy. Hence, in this study, the RF is used as the classifier for the building category classification.

4. EXPERIMENT AND DISCUSSION

To validate the proposed framework, we used 306 aerial images as the study data which includes 73 top-view, 64 forward-view, 47 backward-view, 62 left-view, and 60 right-view images taken by a 5-head Leica RCD30 airborne camera in the experiment. The size of all images is 10336 x 7788 pixels while the four oblique cameras are mounted with a tilt angle of 35 degrees. These images are calibrated by a professional photogrammetric software called Pix4Dmapper which is also used to produce the orthophoto and DSM. The georeferencing accuracy, computed from 9 ground control points, is 2.9 cm and the

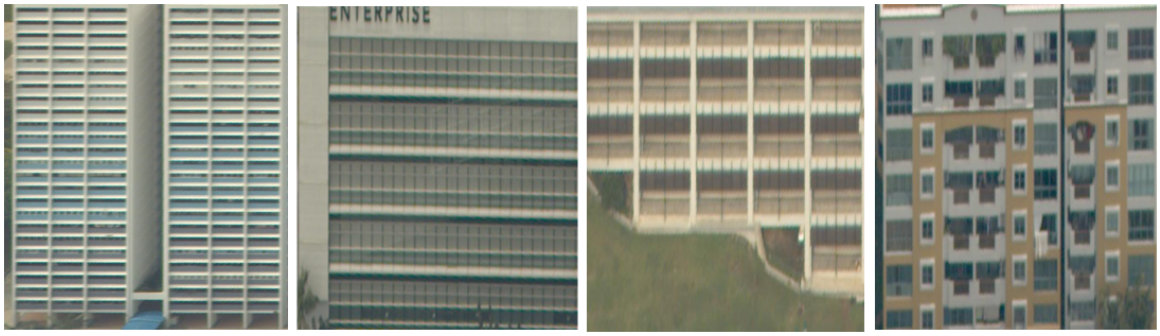

Figure 3. Examples of façade images from oblique aerial images. From left to right are example façade images of residence(regular), education, office, and condominium.

Table 1. The statistics of façade images in different types of buildings.

Type	Num.
Residence(regular)	110
education	78
office	31
condominium	43
total	226

ground sampling distance (GSD) of the orthophoto and DSM is around 7.8 cm. The study area is around the campus of the National University of Singapore (NUS), where contains a variety of buildings. In the experiments, we manually select and draw the boundary of 106 buildings including 1) regular residence; 2) educational building; 3) office; 4) condominium. From these building's boundaries, 262 façade images are cropped, selected, and rectified to front-view to generate the experiment dataset, and the detailed statistics and image examples can be found in Table 1 and Fig. 3, separately. The reader can find more details about the dataset, texture mapping, selection, and rectification for each building in (Xiao et al., 2020). To capture the façade features, average color and the standard deviation in R, G, B channels are calculated, while the Haar-like features are adopted for the texture description. In this experiment, a three-rectangle Haar-like structure is designed and used at the vertical and horizontal direction, separately, at 3 different sizes (total 6 features). Finally, these features are combined to describe the façade textures for their identification. For the random forest classifier, 500 decision trees are used for training, while the number of variables for classification is set as the square root of the feature dimension which is 12 in the experiment. With the different number of training samples, the overall classification accuracy and Kappa value of all façade images (training data is included) are shown in Fig. 4.

From Fig. 4, we can observe both the overall accuracy and Kappa values are increasing with the increased training numbers. Even only use 30% data as training samples, the classification accuracy can still reach 0.6, which is promising in real applications. Since the experiment data is limited (31 office, 43 condominium), the intra-class variability and inter-class similarity can be major challenges for the training on even smaller sub-dataset. However, this result still demonstrated that the façade texture can offer a useful clue for building category classification. With more samples, sophisticated feature extractor, and classifier, e.g., neural networks, we believe the façade image-based building category classification can have much higher accuracy and more practical.

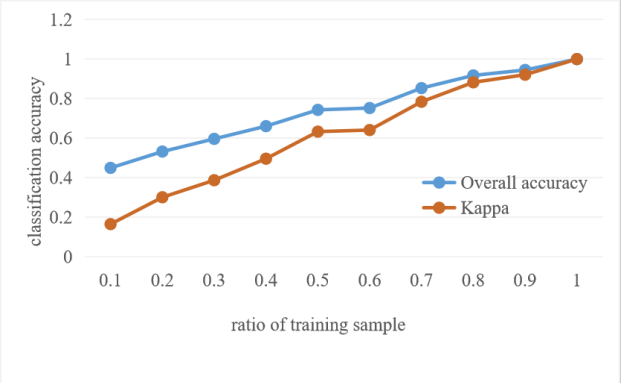

Figure 4. The building category classification results with different numbers of training samples.

ACKNOWLEDGEMENTS

This work is partially finished at Future Cities Laboratory, Singapore-ETH Centre, ETH Zurich. And this material is based on research supported by the National Research Foundation under Virtual Singapore Award No.NRF2015VSG-AA3DCM001-024. Thanks to the Singapore Land Authority (SLA) for the data source.

This research is also supported by the National Natural Science Foundation of China (41701466).

REFERENCES

Christie, G., Fendley, N., Wilson, J., Mukherjee, R., 2017. Functional Map of the World. *arXiv e-prints*, arXiv:1711.07846.

Crow, F. C., 1984. Summed-area tables for texture mapping. *ACM SIGGRAPH computer graphics*, 18number 3, ACM, 207–212.

Deng, X., Newsam, S., 2017. Quantitative Comparison of Open-Source Data for Fine-Grain Mapping of Land Use.

Frueh, C., Sammon, R., Zakhor, A., 2004. Automated texture mapping of 3d city models with oblique aerial imagery. *Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004.*, IEEE, 396–403.

Gruen, A., Schubiger, S., Qin, R., Schrotter, G., Xiong, B., Li, J., Ling, X., Xiao, C., Yao, S., Nuesch, F., 2019. SEMANTICALLY ENRICHED HIGH RESOLUTION LOD 3 BUILDING MODEL GENERATION. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Hartley, R., Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.

Kolbe, T. H., Gröger, G., Plümer, L., 2005. Citygml: Interoperable access to 3d city models. *Geo-information for disaster management*, Springer, 883–899.

Liu, Y., Peng, J., Jiao, L., Liu, Y., 2016. PSOLA: A heuristic land-use allocation model using patch-level operations and knowledge-informed rules. *PloS one*, 11(6).

Minetto, R., Pamplona Segundo, M., Sarkar, S., 2019. Hydra: An Ensemble of Convolutional Neural Networks for Geospatial Land Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 6530-6541.

Newsam, S., 2010. Proximate sensing: Inferring what-is-where from georeferenced photo collections. *IEEE Conference on Computer Vision & Pattern Recognition*.

Rau, J.-Y., Jhan, J.-P., Hsu, Y.-C., 2014. Analysis of oblique aerial images for land cover and point cloud classification in an urban environment. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3), 1304–1319.

Rawat, J., Kumar, M., 2015. Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*, 18(1), 77–84.

Sun, X., Lin, X., Shen, S., Hu, Z., 2017. High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. *ISPRS International Journal of Geo-Information*, 6(8), 245.

Tutzauer, P., Becker, S., Fritsch, D., Niese, T., Deussen, O., 2016. A Study of the Human Comprehension of Building Categories Based on Different 3D Building Representations. *Photogrammetrie-Fernerkundung-Geoinformation*, 2016(5-6), 319–333.

Ty, H., Yang, J., Xuecao, L., Gong, P., 2016. Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sensing*, 8, 151.

Viola, P., Jones, M. et al., 2001. Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, 1(511-518), 3.

Xiao, C., Qin, R., Ling, X., 2020. Urban Land-cover Classification Using Side-View Information from Oblique Images. *Remote Sensing*, 12(3), 390.

Zhang, F., Du, B., Zhang, L., 2015a. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1793–1802.

Zhang, Q., Qin, R., Huang, X., Fang, Y., Liu, L., 2015b. Classification of ultra-high resolution orthophotos combined with DSM using a dual morphological top hat profile. *Remote Sensing*, 7(12), 16422–16440.

Zhu, Y., Deng, X., Newsam, S., 2018. Fine-Grained Land Use Classification at the City Scale Using Ground-Level Images. *IEEE Transactions on Multimedia*, PP.

Zhu, Y., Newsam, S., 2015. Land use classification using convolutional neural networks applied to ground-level images.

Revised May 2020