GEOTAG IMAGE RETRIEVAL FOR SATELLITE IMAGE RECOGNITION

Darshana Charitha Wickramasinghe[a], Tuong-Thuy Vu[a,b]
[a] The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
[b] International University, VNU-HCM, Linh Trung, Thu Duc, HCM city, Vietnam
Email:khgx4dcs@nottingham.edu.my , Tuongthuy.Vu@nottingham.edu.my

KEY WORDS: Volunteered Geographical Data, CBIR, Remote Sensing, Automated Recognition

ABSTRACT: Geotag photos sharing via web-based media and mobile services provide useful information. Geographic metadata and text-based annotations (tags) contained in these volunteered geographical data have been used to enhance the quality of the existing GIS data. However, less attention has been paid to the pictorial information encoded in geotag photo simultaneously. On the other hand, in automated remote sensing image recognition, ground truth information is the essential requirement as the top-view satellite cannot well separate several land use and land cover classes without prior knowledge of the study areas, especially when working with high resolution images.

The key idea of this study is to integrate the volunteered geotag photos and satellite remote sensing data for better automated image recognition. Instead of only using the text annotation conventionally, here we additionally use the pictorial information encoded in geotag photos. Content-Based Image Retrieval (CBIR) approach is adopted for photo interpretation. We experiment with Landsat-8 satellite images and Flickr, Panoramia and GoogleStreetView geotag photos of Klang Valley, Malaysia to identify 4 major land cover class (water, build-up, cropland and other vegetation). The result revealed that the visual information included in geotag photos is a good source for remote sensing image classification. The capacity of volunteered data is not limited to the above 4 land cover class but expandable to the land-use image recognition.

1. INTRODUCTION

Geotag photos come as the volunteered geographical information (VGI) source and its main advantage is free of charge, up to date geospatial information. The famous saying “A picture is worth a thousand words" only reveals the value of visual information contains in a picture, but when capturing a photo, tagged with geographical location; we have additional location information. VGI is commonly used in geospatial application as a supportive information. There are many successful applications like, emergency response, disaster monitoring, land cover classification, risk/suitability analyses. So far, geographical metadata and tagged text contained in the geotag photos have been incorporated in GIS analysis. The visual information, worth thousands more, has not been fully paid attention, because manual photo content annotation is the time consuming and tedious task.

Another advantage of the geotag photo is the ground level view of the Earth surface which provides the detail explanation about the land cover and land use of the tagged location. The satellite remote sensing images are not fully capable to provide such kind of information without ground verification. There is high demand of VGI geotag photo for land use and land cover map verification and validation. If the Geotag photos provide the good interpretation about tagged location, it will be a good ground truth for automated satellite image recognition. The insufficient ground truth and validation data is the main issue in automated mapping.

Hence, the main aim of this study is to retrieve the content land cover and land use types from the geotag photograph prior to proposing an automated satellite image recognition method based on these extracted information.

Related work

(Antoniou et al. 2016) revealed the feasibility of geotag photos as a source of land cover input data. (Leung et al. 2012) in their study used crowdsourced geotag photos for land use mapping. They identified different land use types within the university campuses like academic, sport and residential areas. Geotag photos downloaded from Flickr used for land use interpretation. Bag of visual word method was used for automated photo labelling. Final resolution depends on the density of tagging locations. In (Leung et al. 2014) study, geotag photo was used to classify the developed and undeveloped areas in Great Britain, using both downloaded Flickr and Geograph photos. They considered three different visual features for automated imaged understanding: 1) colour histogram, 2) edge histogram and 3) Gist feature. (Estima et al. 2014) categorised the Flickr geotag photos according to the content land cover and used to land cover mapping. The output map was then compared with the satellite image. (Wegner et al. 2016) used geotag photo to urban tree mapping. Google open street images and satellite images are the cross-reference to extract particular information.

Content based image retrieval (CBIR)

CBIR is a technique for retrieving images on the basis of containing feature in the image. Textual and metadata are the high-level features which explain the content of image. However, manual labelling is time consuming and expensive. Thus, lower level visual features (like colour, texture and shape) are used to describe the image content (Bhad & Komal 2015). CBIR system allows searching and retrieving images that are similar to a given query image. The retrieval performance strongly depends on the utilised feature, called visual description, representing the image content. This section explains the component of CBIR, visual description methods and retrieval methods.

Feature indexing: The initial step is computing visual descriptors that express the images content. Image content can be described by its colour, texture and shape. Normally, these visual components are precomputed and stored in a feature database. There are a number of different visual descriptors, mainly divided into global and local. Global features express image as a whole in terms of colour, texture and shape. The well-known methods are colour histogram, texture histogram, and colour moments.

Local descriptors describe the image patch around the key point. Well-defined points in image are considered as key points. The bunch of key points can be used to express an image; these methods are normally used in object recognition. SIFT and SURF are well known key point description methods and bag of visual word is one approach of object retrieval.

Searching /similarity matching: The main purpose of CBIR is to find the similar image from the collection of images *(predefined content).* The visual descriptors are used to measure the similarity between two images. Initially, all required visual descriptors are calculated and stored in the feature database. Once calculating the corresponding visual descriptors of the query image *(undefined image)*, this descriptor matches with the all the descriptors in the feature database. As a result of the search, a ranked list of images returns to the user. The list is ordered by a degree of similarity. Euclidean distance, cosine similarity, manhattan distance are the commonly used methods of similarity measuring.

2. STUDY AREA and DATA USED

In this study, we explored the capability of geotag photos to support satellite image recognition. We downloaded all the Flickr and Panoramio geotag photos, and the publicly available GoogleStreetview images of that area with their geographical location located within Klang Valley boundary, our study area. The retrieved LULC ground truth data from this study was used to classify the Landsat 8 image of Klang valley. Figure 1 shows the boundary and the photo location distribution across the study area and satellite image. Here, we discriminated the 3 main land cover types 1) vegetation, 2) built-up and 3) water, and 6 other land use types 1) plantation, 2) forest, 3) open area, 4) residential area, 5) commercial area and 6) condominium area.

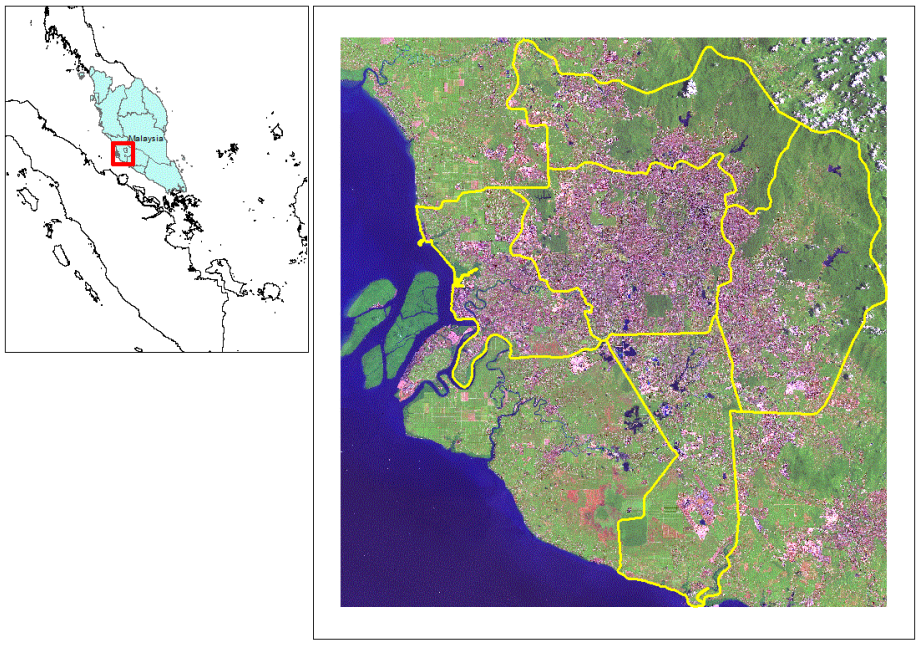

Figure 1. Overview of the study area

3. METHODOLOGY

Automated satellite image recognition framework consists of 2 main stages 1) automated geotag photo annotation, 2) automated satellite image annotation. Figure 2 shows the proposed work flow. Stage I is the ground truth generation and stage II is the satellite image annotation.

3.1 Automated Geotag Photo Annotation

The annotation process is composed of 4 main components, 1) geotag photo database, 2) train/validation photo with LULC label, 3) feature descriptors, 4) SVM classifier. All downloaded geotag photos are arranged inside the photo database with their tagged location. For the training and validation purpose, we manually labelled the 50 sample photos from each LULC types. Similarly, 50 images were selected for the classifier validation. We applied two different methods for land cover detection and land use detection. RGB histogram descriptor matching method is used for land cover labelling, and SIFT features is used to land use labelling.

Land cover labelling

This study expects to categorise the geotag photos into three land cover classes, vegetation, built-up and water features. The colour descriptor is used to distinguish these three classes. We proposed a method based on RGB histogram descriptor for land cover retrieval. We extract the 8-bin Red Green Blue (RGB) histogram for each training photo. This descriptor uses 512 numerical values to represent the colour features of the photo. The histogram is normalized to avoid the conflict of different photo sizes. This histogram descriptor and the land cover labels of training photos are fed as the input of SVM model and the output gave the SVM classifier for the land cover. The model is validated with the validation geotag photo set.

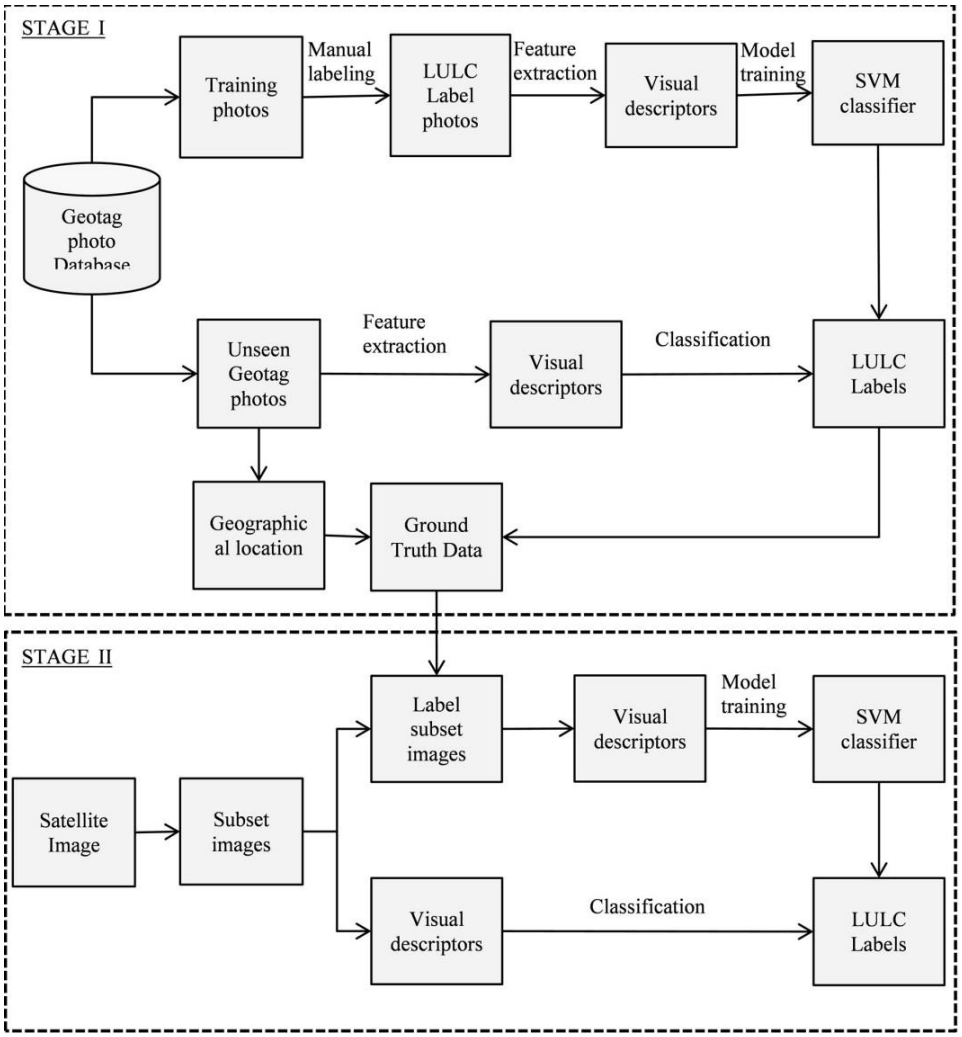

Figure 2. Satellite image recognition frameworks

Land use labelling

The colour descriptor is inappropriate for land use detection because different land use classes may have the same colour distribution (e.g. forest and plantation). Hence, we use the local descriptor instead. We calculate the key point of each training image; 1000 key points, each of which represents the significant feature of the photo, are used to represent the content of a single photo. SIFT feature, consisting of 128 values that represent the single location, is then extracted for each key point location. Subsequently, we use the bag of visual word method to describe the photo content.

The key point description will be the visual words that represent the photo features. In the bag of visual words method, it uses several visual words to describe the content of images. In the training process, it identifies the visual vocabulary (bunch of visual words) of each land use photo. Here we used visual vocabulary with 50 visual words to describe the each land use type. Finally, visual vocabulary and land use label of training data is used to develop land use SVM classifier. The accuracy of classifier is evaluated with the validation photos.

End of the first stage, we create two SVM classifiers for both land cover and land use labelling. First, we separate the unseen (unclassified) geotag photos into three land cover classes. Later, we use land use classifier to separate land use photos. The final classification result will assign the two labels (LC and LU) for each geotag photo.

3.2 Automated satellite image annotation

The goal of the second stage is to incorporate the retrieved ground truth data from the first stage to satellite image recognition. In this study, we experimented with 15m resolution Landsat8 image. The content-based image retrieval approach is suggested for LULC annotation. Due to the lower spatial resolution, in this study we use simple global descriptors for LULC retrieval.

The satellite image annotation process consisted with few steps, 1) image pre-processing 2) image scene generation, 3) LULC training data generation 4) LULC classification.

Image prepossessing

First we select green, infrared and mid-infrared bands (3, 5, 6) for the land cover classification (natural-like colour composite). To obtain better spatial resolution, multispectral image is pan-sharpened with the 15m panchromatic band. Further processing will work on this 15m spatial resolution multispectral Landsat 8 image.

Image scene generation

A satellite image covers a huge area of complex landscape on the Earth surface. In the satellite image retrieval, it is essential to define the boundary of the processing extent, otherwise it will retrieve the content of entire image. In other words, the entire satellite image needs to be divided into small scenes, the extent depends on the expected level of detail, this step is called scene generation. There are various methods of scene generation with predefined GIS data. In this study, we simply divide the whole study area into 20 by 20 pixels (300 m x 300m) grid (Figure 3).

LULC training data generation

Here, we use the labels extracted from geotag photos in stage I as the training data. We assign the LULC label for the scenes which overlap with the photo tagging locations, and extract the RGB histogram descriptor of labelled scene images. The vector of 128x1 dimension is used to represent the visual content of scene satellite image.

LULC classifier

The calculated colour descriptors and corresponding LULC labels are used to create SVM classifier, which is then validated with visually verified locations. Finally, we calculate the RGB histogram for the whole image scenes. Generated SVM classifier assigns the LULC labels for the scene images. The output of this stage is the final LULC annotated map.

4. RESULT

Table 1 shows the numbers of photos downloaded from each platform. There are lots of unrelated photos in Flickr photo dataset such as indoor and personal photos, which were manually removed. Figure 4 depicts the geotag photo distribution over the study area. Google street view photo set get as the training data and select the 50 photos from selected land cover classes (vegetation, build-up, water) and 50 photos form each land use classes. Figure 5 shows the training photo samples of each category.

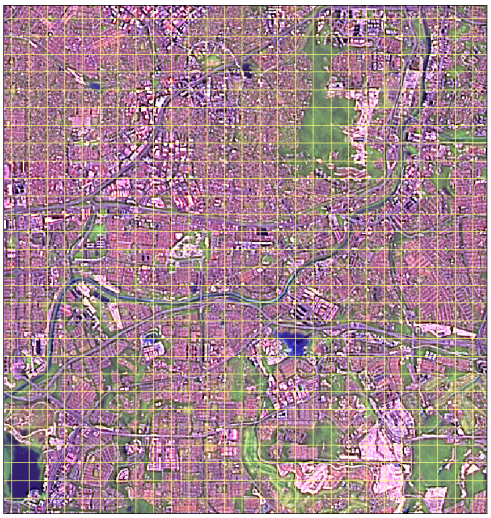

Figure 3 Sample scene images

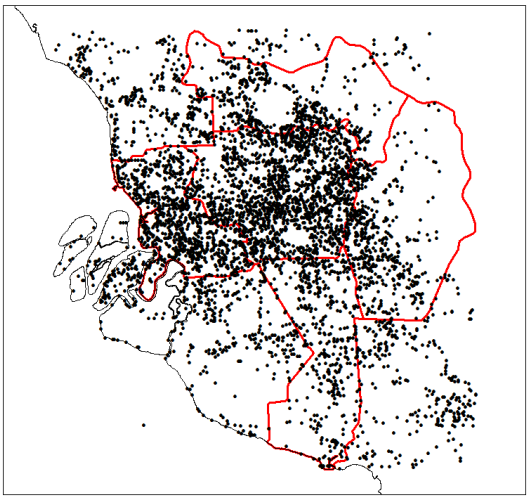

Figure 4 Geo-tag photo distibution

Figure 5 Geotag photo classes a) plantation- palm oil, 2) other free- forest, c) open area-grass, d) residential area, e) commercial area, f) condominium, g) water area

Table 1 Number of geotag photos

Media	Number of photos
Flickr	4206
Panoramio	2000
Google street view	7000

We evaluated the land cover and land use SVM classifiers based on the accuracy, precision and recall values (Table 2-4). Table 2 brief the evaluation result of land cover classifier. Vegetation and build- up classes get 83% and 82% precision value respectively.

We created two land use classifies, one for the vegetation dominant and other or built-up area dominant land. We evaluated both classifiers separately as shown in Tables 3 and 4. The result shows 80% and 79% precisions for forest and palm oil plantation separation whereas the highest value 88% was for open area, because of its light colour and low complexity.

Table 4 presents the accuracy of built-up area dominant land use classification. The classification achieved 73% and 70% precision for residential and commercial areas. Condominium appearance is different from the other two classes and hence it is get high precision.

Table 2 Accuracy assessment- Land cover classifier

		Classified Land cover label			Recall
		Vegetation	Build- up	Water	
Assign Land cover label	Vegetation	45	3	2	90%
	Build- up	4	46	0	92%
	Water	5	7	38	76%
Precision		83%	82%	95%	86%

Table 3 Accuracy assessment- Land use classifier for vegetation dominant class

		Classified Land Use label (Vegetation)			Recall
		Forest	Plantation	Open Area	
Assign Land Use label	Forest	41	7	2	82%
	Plantation	9	37	4	74%
	Open Area	1	3	46	92%
Precision		80%	79%	88%	83%

Table 4 Accuracy assessment- Land use classifier for Build-up area dominant class

		Classified Land Use label (Build-up)			Recall
		Residential	Commercial	Condominium	
Assign Land Use label	Residential	35	15	0	70%
	Commercial	9	39	2	78%
	Condominium	4	2	44	88%
Precision		73%	70%	96%	79%

After evaluation the land cover and land use classifiers, they are used to label the unseen geotag photographs. The photo tagging location consider as the ground truth location. It is used to develop the LULC SVM classifier for the image satellite image classification. Here we evaluated the LULC classifier performance with the manually generated ground truth data. Figure 6 shows the final annotated LULC maps for each land cover class.

Table 5 Accuracy assessment- Satellite image annotation

		LULC Class- manual					Recall
		Forest	Plantation	Open Area	Build-up	Water	
LULC – Class Automated	Forest	43	5	0	1	1	86%
	Plantation	2	45	1	2	0	90%
	Open Area	21	17	1	8	3	2%
	Build-up	4	7	0	39	0	78%
	Water	1	0	2	0	47	94%
Precision		61%	61%	25%	78%	92%	70%

By the photo recognition, we identified the 3 land cover classes and 6 land use classes. But we couldn't distinguish that 6 land use classes, especially 3 build-up land use class (residential, commercial and condominium). However, vegetation based land use classes distinguish from in satellite images but he open area identification not succeed. Because most of open areas located near the build-up area and their extent are considerably small when compare with the image resolution (15m). Hence, the land use recognition is not only dependent on the ground truth availability but also the quality of the satellite image (spatial and spectral resolution).

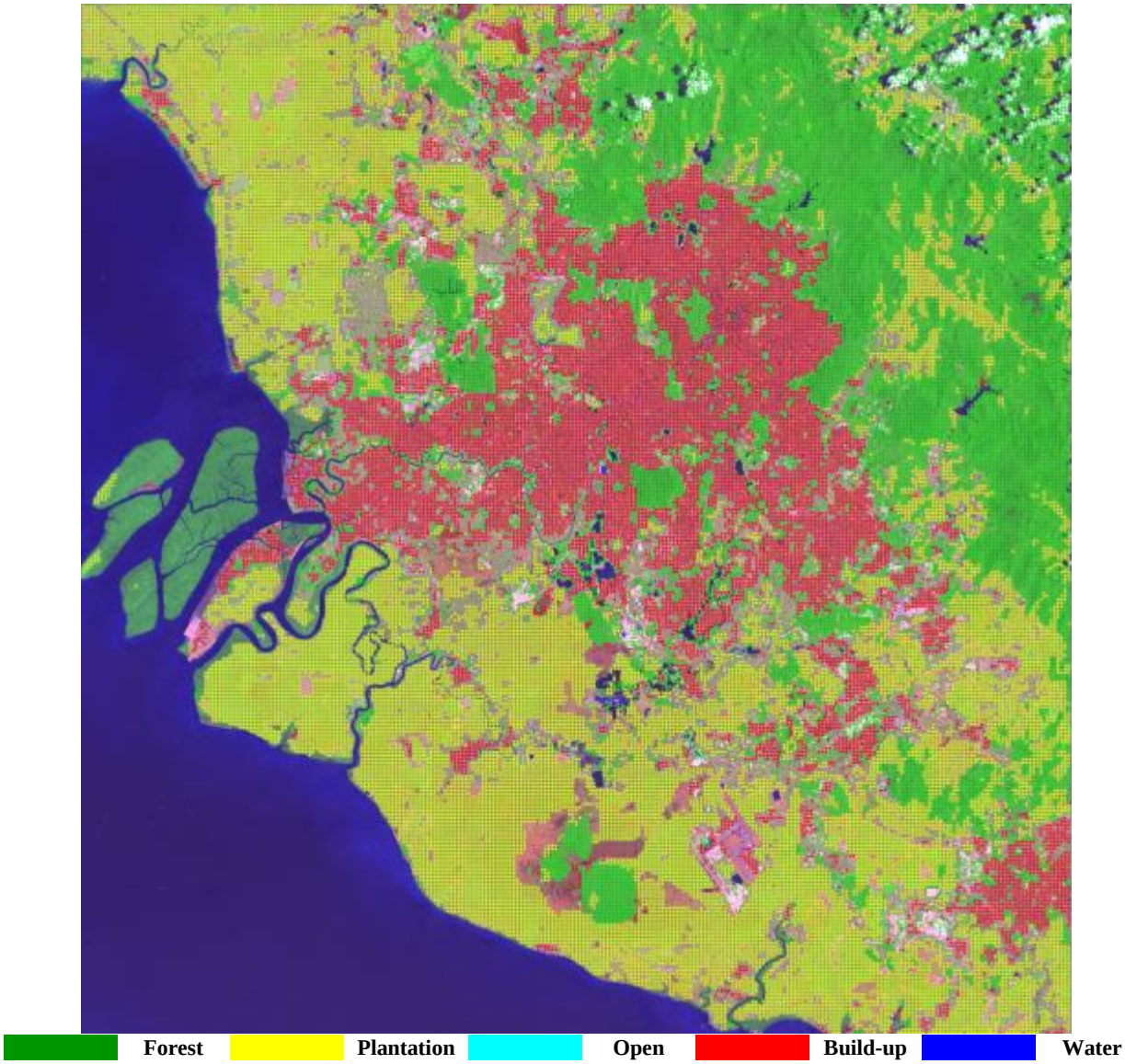

Figure 6 Final LULC annotated satellite image for – open area

5. DISCUSSION

The main goal of this study is to retrieve the content of geotag photographs for the satellite image recognition. The result shows that geotag photos is suitable for the land cover ground truth generation. In land use, the content-based photo retrieval approach could not gain high performance for the several classes. In this study, we focused only on the colour descriptors and SIFT features. Other kinds of image feature descriptors like texture, orientation, pattern features would be useful for further advancement. The result shows that ground truth information was extracted successfully extracted from geotag photos and infers that geotag photos can be used for satellite image recognition.

The tagging location accuracy and precision is the main issue in automated LULC recognition. In some cases, tagging location much differs from the actual location. This problem may be due to manually tagging or the precision of the GPS receivers. The tagging location is acquired from the built-in GPS in mobile devices, of which the positional accuracy varies with the different situations. We cannot expect high positional accuracy from the geotag images. This issue can become severe when working with very high resolution satellite image (<1m spatial resolution).

Photo direction and object distance are the other valuable parameters; these are helpful to increase the positional accuracy of the content of images. Though this information can be found in metadata of photos captured by high tech devices, we cannot expect that it is common in crowdsourcing. It is critical to have a quality assessment framework to enhance the positional accuracy of the geotag photographs.

6. CONCLUSION

Huge number geotag photos are available in the public web servers. These photos provide valuable information about the photo tagging location, and the free and up-to-date visual information of the geographical events. This visual information is a good source for satellite image classification, validation and LULC classification. Adopting existing CBIR methods, we successfully retrieved the LULC features from geotag photos. In a step forward, we exploited the retrieved information from geotag photos for automated satellite image annotation. The outcomes suggested that using geotag photos is a practical, alternative and cost-effective solution to overcome the issue of insufficient ground truth in automated remote sensing image classification.

From the basis of simple integration of crowdsource geotag photo and the remote sensing images in this study, further research will aim to extract the semantic meaning of the geotag photographs and extend to integration of other crowdsourced data like, text, video and maps.

Acknowledgments
This study is part of a project funded by FRGS Malaysia Ministry of Education, grant no. FRGS/2/2013ICT07/UNIM/02/1.

References

Antoniou, V. et al., 2016. Investigating the Feasibility of Geo-Tagged Photographs as Sources of Land Cover Input Data. ISPRS International Journal of Geo-Information, 5(5), p.64.

Bhad, A.V. & Komal, R., 2015. Content based image retrieval a comparative based analysis for feature extraction approach. In 2015 International Conference on Advances in Computer Engineering and Applications. Available at: http://dx.doi.org/10.1109/icacea.2015.7164712.

Estima, J., Fonte, C. & Marco, P., 2014. Comparative study of Land Use/Cover classification using Flickr photos, satellite and Corine Land Cover database. International Conference on Geographical Science Castellon.

Leung, D., Daniel, L. & Shawn, N., 2012. Exploring Geotagged images for land-use classification. In Proceedings of the ACM multimedia 2012 workshop on Geotagging and its applications in multimedia - GeoMM '12. Available at: http://dx.doi.org/10.1145/2390790.2390794.

Leung, D., Daniel, L. & Shawn, N., 2014. Land cover classification using geo-referenced photos. Multimedia tools and applications, 74(24), pp.11741–11761.

Wegner, J. D. et al., 2016. Cataloging Public Objects Using Aerial and Street-Level Images - Urban Trees. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 2016.