Crowd-sourcing Applied to Photograph-Based Automatic Habitat Classification

Mercedes Torres
University Of Nottingham
Jubilee Campus, Wollaton Road
Nottingham, NG8 1BB, UK
psxmt3@nottingham.ac.uk 2nd. author

Dr. Guoping Qiu
The University of Nottingham
&
The University of Nottingham - Ningbo Campus
guoping.qiu@nottingham.edu.cn

ABSTRACT

Habitat classification is a crucial activity for monitoring environmental biodiversity. To date, manual methods, which are laborious, time-consuming and expensive, remain the most successful alternative. Most automatic methods use remote-sensed imagery but remotely sensed images lack the necessary level of detail. Previous studies have treated automatic habitat classification as an image-annotation problem and have developed a framework that uses ground-taken photographs, feature extraction and a random-forest-based classifier to automatically annotate unseen photographs with their habitats. This paper builds on this previous framework with two new contributions that explore the benefits of applying crowd-sourcing methodologies to automatically collect, annotate and classify habitats. First, we use Geograph, a crowd-sourcing photograph website, to collect a larger geo-referenced ground-taken photograph database, with over 3,000 photographs and 11,000 habitats. We tested the original framework on this much larger database and show that it maintains its success rate. Second, we use a crowd-sourcing mechanism to obtain higher-level semantic features, designed to improve the limitations that visual features have for Fine-Grained Visual Categorization (FGVC) problems, such as habitat classification. Results show that the inclusion of these features improves the performance of a previous framework, particularly in terms of precision.

Categories and Subject Descriptors

H.3.4 [**Database management**]: Database applications - Image databases

General Terms

Algorithms, Design, Experimentation.

Keywords

Image annotation; image classification; crowd-sourcing; habitat classification; FGVC

MAED'14, November 7, 2014, Orlando, USA.

http://dx.doi.org/10.1145/2661821.2661824.

1. INTRODUCTION

Habitat classification is the process of mapping all habitats present in an area according to a determined scheme [9]. The purpose of classifying habitats is twofold: it helps to reduce the complexity present in the natural world and, by categorizing habitats, their characterization and comparison can be done much more efficiently and effectively. In the UK, Phase 1 is one of the most widely used schemes [9]. This standardized hierarchical classification was designed to provide a detailed record of the vegetation present in an area. However, it relies very heavily on human surveyors. This is laborious, expensive, time consuming and, given the similarities between some of the habitat classes, subjective [9].

Approaches have been developed with the aim of automating the habitat classification process but, to our knowledge, the only alternative which takes into consideration the whole of the Phase 1 scheme is presented in [19, 20, 18]. One of the main reasons why no other alternatives with accurate results have been developed is because most of the automatic habitat classification methods for other schemes use remote-sensed imagery [4]. However, given the level of detail necessary to distinguish between some Phase 1 habitats, remote-sensed imagery has been proven to be insufficient [17]. We approach habitat classification as a FGVC problem and we develop an automatic image annotation framework based on feature extraction and random projection forests to automatically classify habitats. Moreover, we have used an alternative source of information that has obtained promising results, as shown in [19, 18]: ground-taken imagery. These photographs present two advantages over remote-sensed imagery. First, ground-taken photography has a greater degree of detail. For FGVC problems, such as habitat classification, this is a decisive trait, since details will be crucial to differentiate between similar habitat classes, i.e. different types of grasses or heath mosaics. Second, they can be obtained more easily than remote-sensed imagery, since the only requirement are digital photographs.

In this paper, we present a further development of the framework presented in [19]. In particular, we study the effect on our system of crowd-sourcing methods when incorporated to two components: the source data and the features extracted. We make two contributions. First, we have updated our database using Geograph, a crow-sourcing website whose aim is to collect geographically representative photographs and information for every square kilometre of Great Britain and Ireland [14]. By using this crowd-sourcing site, we have benefited from their large collection

of photographs to create a vast and robust database in a straightforward manner. This database, called Habitat 3K, will be made publicly available for the research community. Secondly, we have used a crowd-sourcing methodology to obtain higher-level semantic features using a "Wisdom-of-the-crowd" approach. Several users are asked to annotate Habitat 3K photographs with a simpler set of classes. Their answers, and the certainty levels on their annotations are recorded and transformed into medium-level features, which we combine with low-level features as the input of our classifier. We have carried out extensive experiments to test the influence of these two elements and recall and precision results have shown that our original Random-Projection Forest design is stable enough to handle a larger database, such as Habitat 3K, which is three times larger than the original Habitat 1K database presented in [18]. Experiments also show that the inclusion of semantic information greatly benefits the classification of all classes, particularly the precision of those which are commonly harder to classify, such as Tall Herb and Fern (C) and Heathland (D). Consequently, we can conclude that the addition of crowd-sourcing mechanisms to the task of two automatic habitat classification has had positive effects on the performance of the framework.

2. PREVIOUS WORK

Habitat Classification: There are numerous habitat classification schemes that have been developed worldwide [9]. Although their objectives and parameters are quite different, the classifications with better results rely heavily on manual classification. This is labour intensive, costly, subjective and can take a significant amount of time. On the other hand, most of the automatic approaches proposed either develop their own schemes [8] or focus on classifying particular standard habitats [23]. The former leads to results which are very dependent on the site and not easily comparable with other schemes and the later leads to relative or incomplete results [1]. Most of the automatic approaches proposed in the literature use remote sensing imagery [8] in their design, which are particularly unsuited for Phase 1 classification [17]. The use of aerial and satellite imagery to categorize Phase 1 habitats presents several disadvantages. The most crucial disadvantage is their lack of detail. This results in incomplete data and coverage, little or non-existent species information and even low-quality information, caused by the presence of clouds or intensity differences within the raster photographs. In this paper, we study the performance of ground-taken photographs, which are much easier to obtain than remote-sensed imagery and present a much finer level of detail, which can be crucial when classifying FGVC problems.

Image Processing: Automatic Phase 1 habitat classification using ground-taken photos can be approached as an image annotation problem. The aim is to identify which habitats are present in which photos and where they are localized. There are many approaches that have been developed for image annotation with general classes. [13] combined image annotation with semantic information and bag-of-features to classify photographs according to classes such as grass, water, chair, road and cat. [15] used semantic texton forests to annotate and classify images with a similar scheme. [11] also developed a method for scene recognition based on partitioning an image into increasingly finer subregions and computing their histograms. However, what makes the problem of habitat classification different from general annotation problems is the nature of the classes. Instead of conventional and clearly separable classes, such as *trees, grass, boat, water* [13], Phase 1 is a hierarchical classification whose classes are difficult to identify even for human surveyors. The aim, instead of classifying trees or water, is to classify *which* kind of trees (broad-leaved or coniferous) or water (standing or running) appear in the photographs. In Computer Vision, this type of problem is commonly referred to as fine-grained visual categorization problems [5]. Other examples include the categorization of leaves [10] and birds [2]. FGVC and image annotation are deeply connected, as most FGVC datasets and approaches work with different types of annotations. For example CUB-200-2011 is a dataset for birds with parts and attributes [21]. Additionall,y feature selection is crucial for FGVC problems and can determine the success or failure of the classification. We propose the inclusion of higher-level features to automatic habitat classification to improve the limitations that using visual features entail when classifying visually similar classes. Higher-level features are designed to incorporate semantic information from an image that low-level features are unable to collect. Crowd-sourcing methodologies can be used to extract this knowledge [7]. Through a "Wisdom-of-many" approach [16], the opinions of several non-expert people are taken into account in the classification. The combination of low-level and medium-level features is designed to help classify habitats which share very similar visual properties and improve accuracy of our framework as a whole.

3. PHASE 1 HABITAT CLASSIFICATION

We are using Phase 1 Habitat Classification because it is one of the most widely-used schemes. A robust classification scheme, such as Phase 1, is an essential tool for nature conservation since being able to identify and record species, ecological communities and habitat types is vital to ensure their protection. Phase 1 is a standardized hierarchical system for classifying and mapping wildlife habitats. It was first devised in the 1970s in the UK and it is designed for rapid wildlife mapping over large areas of countryside. It comprises ten broad categories: Woodland and scrub (A), Grassland and marsh (B), Tall herb and fern (C), Heathland (D), Mires (E), Swamp (F), Open water (G), Coastland (H), Rock exposure (I), Miscellaneous (J). In total, the Phase 1 classification scheme contains 155 recognized habitat types organized in three different tiers, from more general to more specific. Each class is identified by its name, an alphanumeric code, a description and a mapping color. Current implementations of the Phase 1 scheme rely on human surveyors to map the habitats. This has many disadvantages: surveyors need to be trained specifically in Phase 1 classification; depending on the size of the site to audit, manual habitat classification can be expensive and time consuming; finally, given the degree of detail required, it can be extremely laborious.

4. HABITAT CLASSIFICATION AS AN IMAGE ANNOTATION PROBLEM

We approach automatic habitat classification from an image annotation perspective. The input are ground-taken photographs and the output is a list of all possible habitats ranked from most probable to less probable. Figure 1

shows an overview of the whole system. The method has two main steps: first, in order to work with the images in a more efficient manner, image features are extracted. Second, the features extracted are used as the input of a Random Projection Forest [19, 18] which calculates the probability of occurrence of each possible habitat in the ground-taken photograph. In this paper, we have focused on studying the effect of crowd-sourcing applied to the first step of the framework. For the second step, we follow the design presented in [19].

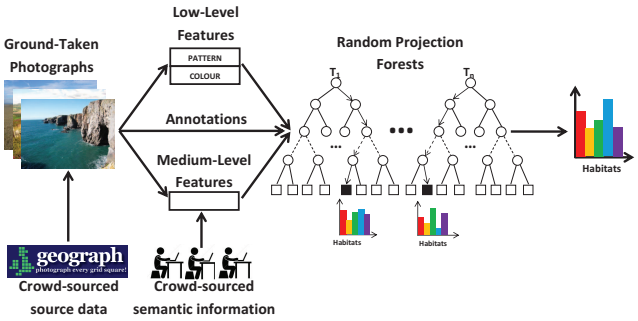

Figure 1: Image Annotation-Based Habitat Classification.

4.1 Ground-Taken Photographs: Habitat 3K

In [18], Torres and Qiu presented a 1,086 ground-taken photograph database, called Habitat 1K (H1K). In an effort to include more variability on the photographs conditions and to increase the number of habitats in the database we created Habitat 3K (H3K). This database will be made publicly available. H3K was created combining the photographs from Habitat 1K and 2,005 photographs from Geograph [14]. Geograph's aim is to collect, publish, organise and preserve representative images and associated information for every square kilometre of Great Britain and Ireland [14]. It maintains a database of over four million ground-taken photographs and it also stores their associated metadata, such as geographical location, the time of the photo, etc. These data are uploaded by a multitude of users, spread all over the UK and they are freely available to the public. Geograph photographs can be tagged and annotated.

We used Geograph to collect 2005 additional photographs using their search-by-tag feature and by searching for the ground-taken photographs with any of these tags: Arable, Boundary, Coastal, Flat landscapes, Grassland, Heath, Scrub, Hedge, Lakes, Park and Public Gardens, Rivers, Streams, Drainage, Rocks, Scree, Cliffs, Wall, Woodland, Forest. This feature enabled us not only to increase rapidly the size of our database but also to access photographs from habitats that, given our current geographical location, were difficult to access, such as a wide range of Coastland (H) habitats. In comparison with H1K, H3K is three times its size, 3094 against 1086 photographs, and it contains more than twice the number of habitats, 11,344 habitat instances against 4,344. Photographs from this database are shown in Figure 2. H3K contains habitats from all possible Phase 1 classes except E and F. H3K has an average of 3.66 annotations per image, a minimum number of 1 annotation per image and a maximum of 6. Additionally, it contains a mixture of high and low resolution photographs, taken during all twelve months of the year in Great Britain. Its classification is a mixture of the classification done by an expert in Phase 1 and the classification obtained from Geograph's tagging system.

Figure 2: Habitat 3K.

4.2 Higher-Level Features

Low-level features, such as the pattern features that were extracted in [19], commonly collect only visual information in the form of global or local statistics. However, there are objects that, while belonging to completely different classes, might have similar visual properties. This is particularly prominent in FGVC problems and makes their automatic classification process extremely complicated if only visual features are taken into consideration. For example, based on colour, texture or pattern features alone, it is impossible to distinguish a tree that belongs to a Woodland (A.1) habitat or a tree that belongs to a Hedge and Trees (J.1.2.) formation. In these cases, there is a clear gap between the visual characteristics of the objets within a photograph and their semantic meaning. This phenomenon is known in the Computer Vision field as the "Semantic Gap" [3]. It is crucial to notice that the semantic gap problem is even more pronounced and has more effect in FGVC problems, such automatic habitat classification. FGVC problems aim to accurately classify between classes that are visually similar and have similar semantics [22]. These classes, as shown in Figure 3, can be indistinguishable to the untrained eye. In an effort to bridge this gap, we propose the introduction of semantic information in the classification process. For this, a new type of feature, often referred to higher-level features, has been proposed [6].

We employ a crowd-sourcing methodology to extract semantic information in an effort to improve the classification. We refer to this semantic information as medium-level knowledge. From them, we create medium-level features. The aim of using humans and crowd-sourcing to collect this semantic information is to create a system that can benefit from both humans' strengths, such as being able to differentiate between different classes just by looking at a photograph, and computers' strengths, such as being able to carry out complicated calculations at a fast speed. Moreover, in order to take into consideration visual and semantic

(a) S.I. Grassland (b) I. Grassland

Figure 3: Visual Similarity in FGVC. These similar photos belong to different classes: Semi-Improved and Improved Grassland.

information during the classification of habitats, we combine low-level and medium-level features. The combination of low-level and medium-level features is designed to help classify habitats which share very similar visual properties and improve accuracy of our framework as a whole.

To create the medium-level features, we show users photographs from our Habitat 3K database and we ask them to annotate them with the classes that they see. Instead of using the whole Phase 1 scheme, which can be confusing for untrained humans, we have created a set of thirty two basic classes, shown in Table 1, that users use to annotate the photographs. Moreover, we ask them to record the certainty of their classification. We then use the confidence measures collected with the annotations from each user. For each image x in the database, all the users' responses stored in a 32-dimension feature vector $H(x) = (h_1, h_2, , h_{23})$ that is generated as follows:

$$h_i = \begin{cases} c_i & \text{if class } i \text{ is present} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where c_i is the degree of confidence that the user has in that their annotation belongs to the correct class i in the photograph x. Consequently, the vector H is what we will refer to as medium-level features. This feature vector is then combined with low-level visual features and used as the input of the random-projection-forest classifier in our framework to automatically annotate unseen ground-taken photographs, as shown in Figure 1.

Table 1: Medium-level Annotations.

Annotations		
Trees - leaves	Bushes	Reed
Trees - mixed	Fern	Herbs
Trees - no leaves	Grass - flowers	Grass - uniform
Grass - green	Heath	Water - running
Water - still	Cliff - near water	Cliff - no water
Rocks - Large	Rocks - Small	Sand
Shingles	Crops	Wall
Fence	Hedge	Sky
Other	Blue	Green
Red	Yellow	White
Brown	Winter	Summer

5. EXPERIMENTS

We designed two experiments to test each of the contributions of this paper. First, we evaluated the stability of our original framework by comparing precision and recall metrics [19] when using both H1K and H3K as input. As shown in Figure 1, we extract two types of low-level features: pattern features (CPAM [12]) and colour features (colour histograms and colour moments). Second, in order to evaluate the use of higher-level features, we compare the results obtained in the previous experiment with those obtained from combining medium and low-level features when classifying H3K habitats. Additionally, in order to understand better the effect of semantic features, we have calculated the confusion matrices for first-tier habitats in H3K.

6. RESULTS

Figure 4 shows the recall and precision [19] results for first-tier Phase 1 habitat classification when using our framework with H1K and H3K with colour and pattern features and with and without medium-level features. Moreover, Table 2 shows the confusion matrix for H3K when higher-level features were excluded and included in our framework.

As can be seen when comparing H1K and H3K performances, our framework is stable. It is able to maintain the same level of accuracy in all cases, and, in the case of Heathland (D) habitats, even obtains more accurate precision results. Their classification in general improves greatly in H3K given their much larger number of instances, 824 in H3K against 135 in H1K. Moreover, it can also be seen that the introduction of medium-level features helps the classification of all habitats. Table 2 shows that the inclusion of semantic information increases the percentages of correctly classified habitats, presented in the diagonal of the matrix. Particularly, Tall herb and Fern (C) classification shows an increase of over 10% in accuracy, as do Heathland (D), Open Water (G) and Coastland (H) habitats. Previously, these habitats obtained less accurate results because the share similar visual characteristics with other habitats. For example, heathland mosaics are extremely similar to scrub. Since in our original framework the only information that was extracted was their visual properties, we were unable to distinguish properly between them. However, by using crowd-sourcing methods and employing humans, we were able to add semantic information in the classification process which was crucial to increase their successful classification rate.

7. CONCLUSIONS AND FURTHER WORK

Automatic Phase 1 habitat classification is a FGVC problem with many ecological applications. We built on previous work [19, 18, 20], which presented a automatic image annotation framework for habitat classification using ground-taken photographs. We study the effect of crowd-sourcing methods applied to habitat classification and make two contributions: a 3,000 fully-annotated publicly-available database, Habitat 3K, collected using the crowd-sourcing database Geograph, and the creation and collection of semantic higher-level features through a "Wisdom-of-the-crowd" approach in which users annotate photographs with a simpler set of classes. These annotations are combined with low-level visual features as the input of our Random Projection Classifier. Experiments show that crowd-sourcing mechanisms are a beneficial addition to our framework and that they increase recall and precision of even the most difficult habitats to classify, such as Tall Herb and Fern (C) and Heathland (D). Further work will include the addition of

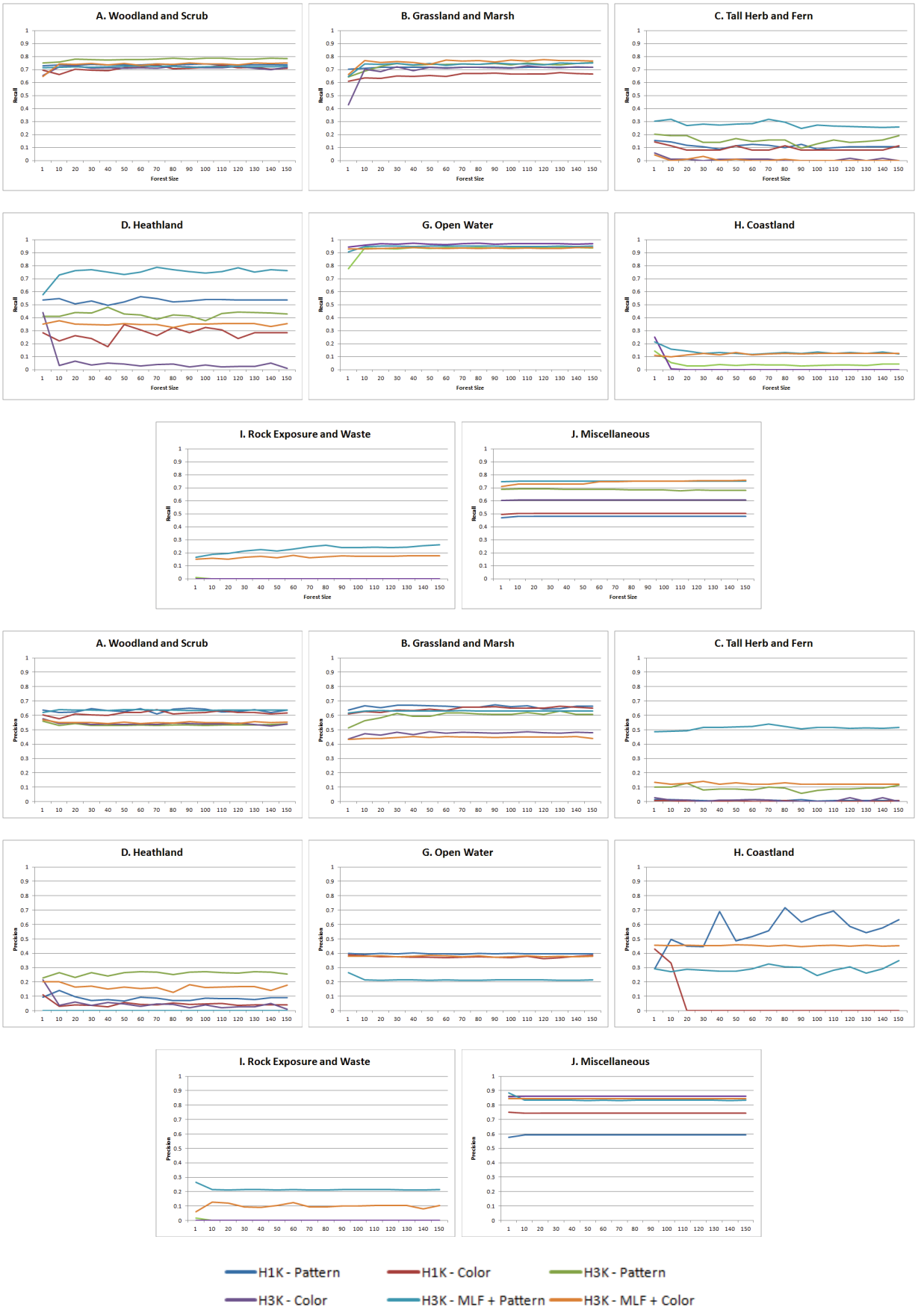

Figure 4: First-tier recall (first three rows) and precision (last three rows) results for experiments with pattern and colour features with H1K and H3K.

Table 2: Confusion Matrix of H3K when we do not use higher-level features (left) and when we do (right).

	A	B	C	D	G	H	I	J
A	**52.92/57.82**	9.99	17.83/13.51	13.73	0	0	0	5.53/4.95
B	5.10/4.89	**52.14/58.11**	5.35/4.63	18.37/15.8	0	0	0	19.04/16.57
C	42.57/33.99	6.60/5.61	**3.63/13.86**	34.32	0	0	0	12.87/12.21
D	29.98/27.55	43.81/37.74	0.97/0.24	**5.83/20.63**	0	0	4/2	15.41/11.41
G	0	3/2	0	4/3	**35.55/43.08**	0	22/18	35.66/33.55
H	6/4	11/10	0	0	0	**3/20**	37/28	43/38
I	8/7	0	0	3	0	67/59	**0/9**	22
J	4.11/3.40	2.66/2.25	0.53/0.32	1.01/0.85	0	0	0.16/0.05	**91.53/93.14**

more photographs to our database, further testing with other types of low-level visual features and more medium-level annotations for the users to employ.

8. ACKNOWLEDGMENTS

Mercedes Torres is supported by the Horizon DTC at the University of Nottingham (RCUK Grant No. EP/G037574/1). Guoping Qiu is supported by the IDIC at the UNNC and by Ningbo Science and Technology Bureau projects 2013D10008 and 2012B10055.

9. REFERENCES

[1] R. Alexander, A. Millington, et al. *Vegetation mapping: From patch to planet.* John Wiley & Sons, 2000.

[2] T. Berg, J. Liu, S. W. Lee, M. L. Alexander, D. W. Jacobs, and P. N. Belhumeur. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[3] C. Bishop et al. *Pattern recognition and machine learning*, volume 1. Springer New York, 2006.

[4] D. Boyd, C. Sanchez-Hernandez, and G. Foody. Mapping a specific class for priority habitats monitoring from satellite sensor data. *IJRS*, 27(13):2631–2644, 2006.

[5] S. Branson, G. Van Horn, C. Wah, P. Perona, and S. Belongie. The ignorant led by the blind: A hybrid human-machine vision system for fine-grained categorization. *IJCV'14*, pages 1–27, 2014.

[6] S. Chang, W. Hsu, L. Kennedy, L. Xie, A. Yanagawa, E. Zavesky, and D. Zhang. Columbia university trecvid-2005 video search and high-level feature extraction. In *NIST TRECVID workshop, Gaithersburg, MD*, 2005.

[7] J. Deng, J. Krause, and F. Li. Fine-grained crowdsourcing for fine-grained recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 580–587, June 2013.

[8] R. A. Díaz Varela, P. Ramil Rego, S. Calvo Iglesias, and C. Muñoz Sobrino. Automatic habitat classification methods based on satellite images: A practical assessment in the nw iberia coastal mountains. *Environmental Monitoring and Assessment*, 144(1-3):229–250, 2008.

[9] Joint Nature Conservation Committee. Handbook for Phase 1 habitat survey - a technique for environmental audit, 2010.

[10] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *ECCV'12*, pages 502–516, 2012.

[11] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR'06*, volume 2, pages 2169–2178, 2006.

[12] G. Qiu. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition*, 35(8):1675–1686, 2002.

[13] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *ICCV'07*, pages 14–21, 2007.

[14] G. Rogers. Geograph, http://www.geograph.org.uk, 2005.

[15] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR'08*, 2008.

[16] J. Surowiecki. *The wisdom of crowds*. Random House LLC, 2005.

[17] M. Torres. Automatic habitat classification using aerial imagery. In *GIS Research UK 20th Annual Conference (GISRUK)*, volume 1, pages 11–13, 2012.

[18] M. Torres and Q. Guoping. Habitat classification using random forest based image annotation. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 1491–1495, Sept 2013.

[19] M. Torres and G. Qiu. Grass, scrub, trees and random forest. In *MAED 2012 - 2012 ACM Workshop on Multimedia Analysis for Ecological Data, ACMM'12*, pages 1–6, 2012.

[20] M. Torres and G. Qiu. Automatic habitat classification using image analysis and random forest. *Ecological Informatics*, 2013.

[21] C. Wah, S. Branson, P. Perona, and S. Belongie. Multiclass recognition and part localization with humans in the loop. In *ICCV'11*, pages 2524–2531, 2011.

[22] L. Xie, Q. Tian, S. Yan, and B. Zhang. Hierarchical part matching for fine-grained visual categorization. Technical report, Technical Report, Department of Computer Science and Technology, Tsinghua Univerity, 2013.

[23] C. Zhang and Z. Xie. Combining object-based texture measures with a neural network for vegetation mapping in the everglades from hyperspectral imagery. *Remote Sensing of Environment*, 124:310–320, 2012.