

# Automatic Habitat Classification Using Aerial Imagery

Mercedes Torres<sup>1</sup>

<sup>1</sup>Horizon Doctoral Training Centre, School of Computer Science  
The University of Nottingham  
Wollaton Road  
Nottingham  
NG8 1BB  
psxmt3@nottingham.ac.uk

**Summary:** Manual habitat classification is labour intensive, costly, subjective and time consuming. This paper presents an automatic habitat classification method for aerial photography using SIFT descriptors and BOVW and studies its recall ability and its accuracy in a retrieval and a classification scenario, respectively.

**KEYWORDS:** Habitat classification, image processing, aerial imagery, SIFT descriptors, bag of visual words.

## 1. Introduction

Habitat classification and its applications (e.g. habitat monitoring, identification of rare species, etcetera) are important challenges researched by environmental bodies and mapping agencies. However, manual habitat classification is labour intensive, costly, subjective and time consuming (Chen and Rau, 1997).

From an image processing perspective, habitat classification can be achieved using two different approaches: a retrieval approach, whose objective is to retrieve photos from the same habitat as the query, and a classification approach, whose objective is to correctly classify the query image using photos from a database. In this paper, a content-based approach based on feature extraction from aerial imagery is described and its performance in these two scenarios is evaluated.

## 2. Application to Habitat Classification

This paper expands work previously done by Sivic and Zisserman (2003) in which visual words were extracted to describe video frames and to detect and retrieve objects under varying conditions. Visual words are used because they enable us to describe images using only a numerical vector, an inverse frequency vector. Consequently, the complicated task of comparing images is reduced to calculating the distances between their respective frequency vectors. To obtain those inverse frequency vectors, a codebook, along with the visual words of each image are needed.

A codebook is a glossary of the most descriptive visual words, called in this case code words. For this project, a 100-code-word codebook has been calculated using k-means clustering and the Corel Database. This database reunited two important requisites necessary to generate the codebook: it is varied, so the resulting code words will be descriptive, and independent of the testing images, so the same codebook can be used with different testing sets.

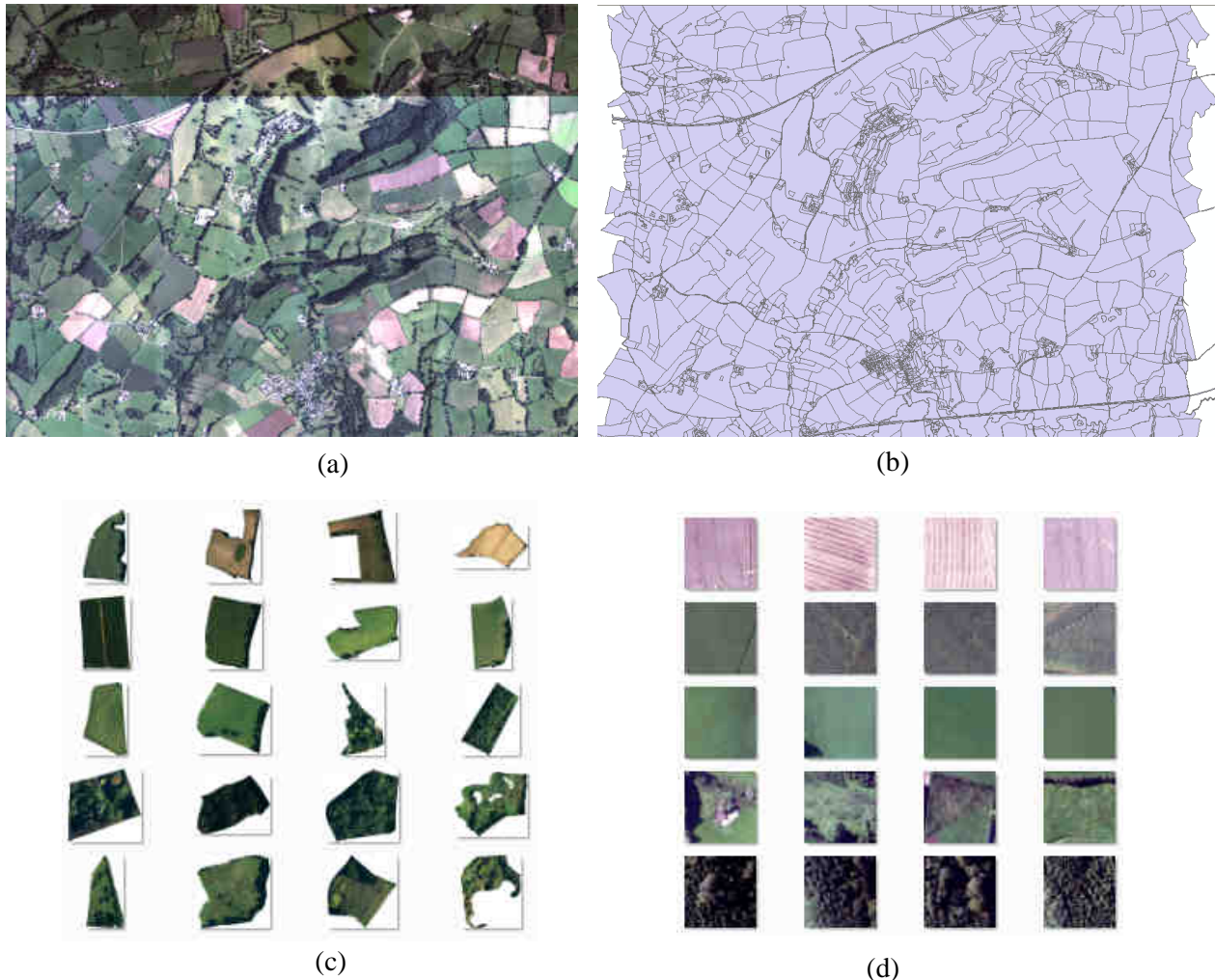
On the other hand, given the varied nature of the aerial photography, the visual words extracted are Scale-Invariant-Feature-Transform (SIFT) descriptors. These descriptors are suitable candidates to describe images because they detect lighting-, perspective-, orientation - and scale-invariant regions. Each image will have a variable number of visual words.

The inverse frequency vector describing each aerial image is generated by measuring the frequency of appearance of the code words in relation to its own visual words (Sivic and Zisserman, 2003). By using the inverse frequency, visual words that appear less will have more weight when describing the images.

## 2.1. Data

Figure 1 shows the data involved:

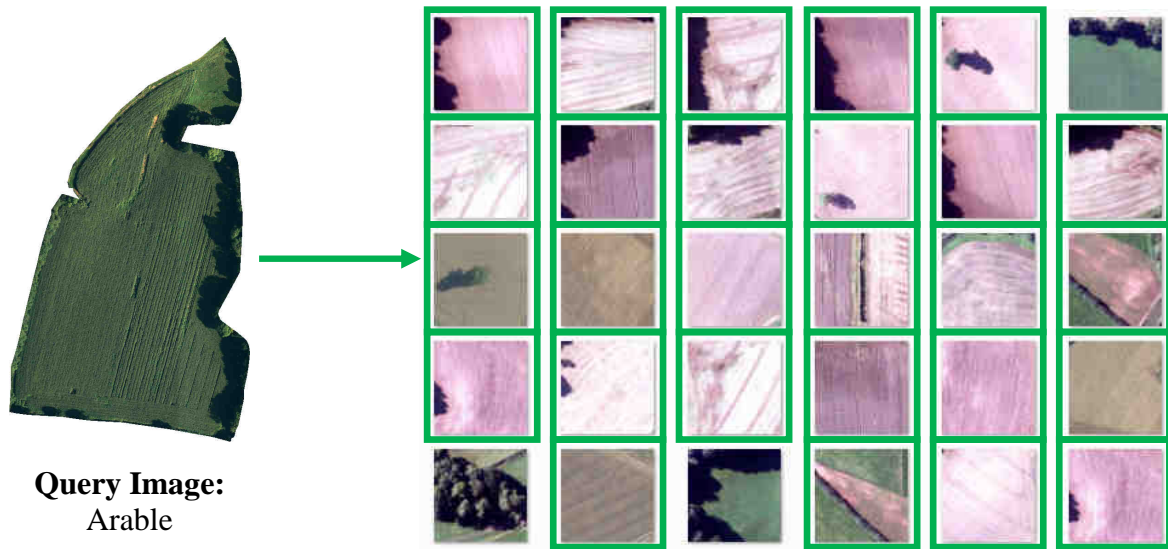
1. Raster image: aerial photograph composed by a variable number of plots with different lighting conditions. Instead of using the whole image in the query and then using a spatial extension in the retrieval process (Yang and Newsam 2010), OS MasterMap was used to clip the images.
2. Query set: all the clipped images obtained from the raster and classified by an expert.
3. Test set: ground-truth catalogue classified by an expert in Phase 1 Habitat Survey (JNCC, 2010) with a large number of images that represent each different habitat class.



**Figure 1.** (a) Raster image, (b) OS MasterMap with polygon information, (c) clipped 3-channel images and (d) Test set classified by an expert.

## 2.2. Retrieval

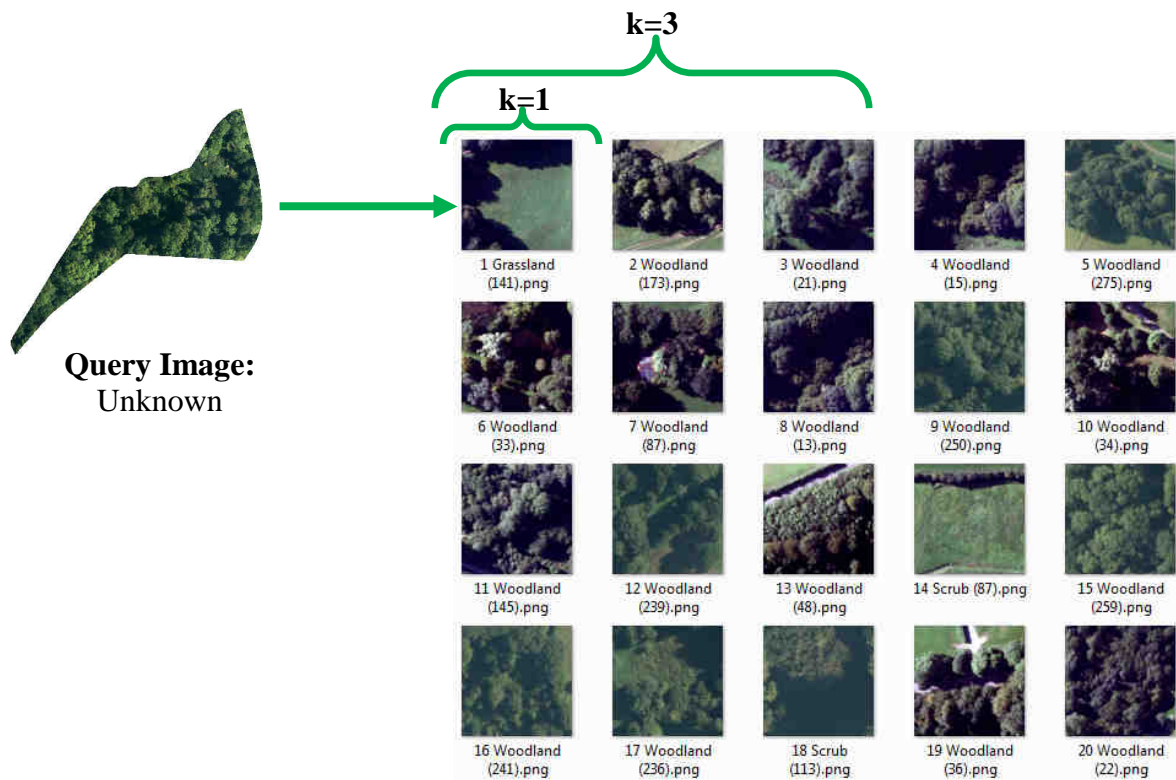
In this case, as shown in Figure 2, the habitat class of the query image is known. The objective is to retrieve all the photos from the query set that belong to the same category as the query image. This is done by calculating the Euclidean distance between the frequency vectors that describe the query image and the images in the test set and indexing the results.



**Figure 2.** Retrieval. Using the query image, we are able to retrieve 27 arable habitats (outlined in bold) within the first 30 results.

### 2.3. Classification

In this case, as shown in Figure 3, the class of the query image is unknown. The objective is to classify it using its closest images in the test set. K-NN (Cover and Hart, 1967) is used to decide the class of the query image by averaging the k first results.



**Figure 3.** Classification. Using k-NN, the query image is classified by averaging the k first results. For k=1, the query image would be classified as “Grassland”. However, for k=3 or larger it would be correctly classified as “Woodland”.

## 4. Results

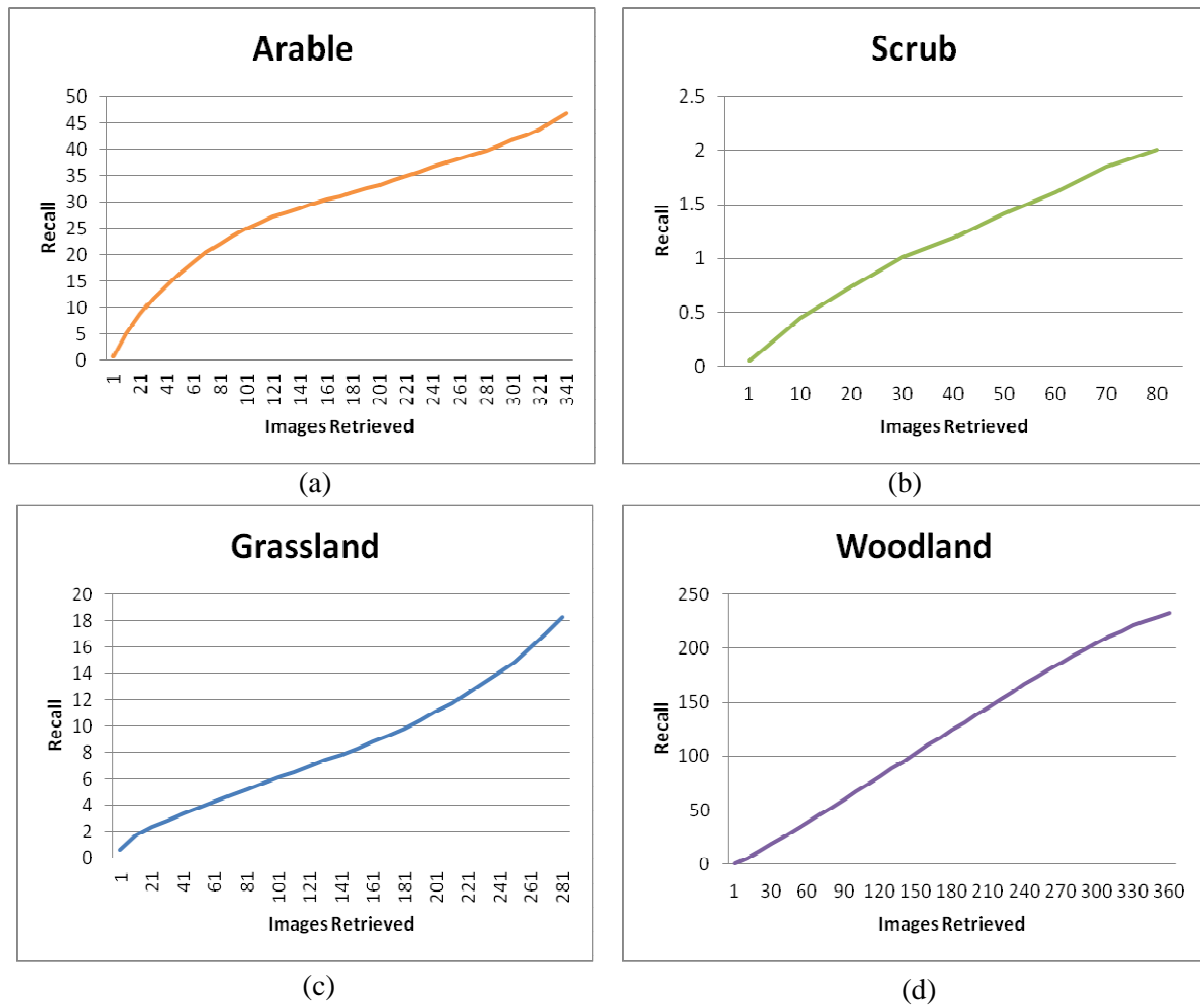
To test the two scenarios, imagery from two different locations, a query area and a test area, were classified by an expert. Table 1 shows the number of images corresponding to the four habitats retrieved and classified in both areas.

**Table 1.** Number of images for each habitat extracted from the query and the test area.

<i>Habitat</i>	<i>Query Area</i>	<i>Test Area</i>
Arable	68	346
Grassland	411	285
Scrub	12	80
Woodland	259	361

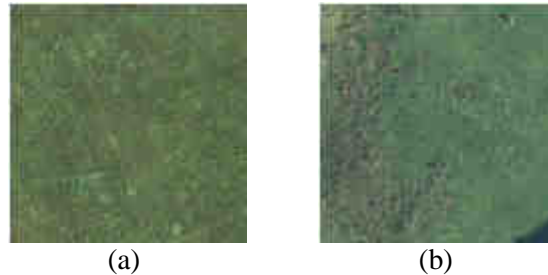
### 4.1. Retrieval

The retrieval accuracy of the approach, shown in Figure 4, was measured by calculating its recall ability. By varying the number of retrieved images from one to the number of images of that habitat class in the test set, an average of the number of correct answer retrieved was calculated.



**Figure 4.** Recall of (a) Arable, (b) Scrub, (c) Grassland and (d) Woodland images. Perfect recall ability would imply that all the images retrieved belong to the same class as the query image.

Results show that as the number of images retrieved increase, so does the recall, which is consistent with the approach followed. Recall results concerning grassland and scrub are significantly low. This is mainly due to the fact that scrub and grassland habitats can have similar intensity properties and, consequently, the visual words extracted from the images can be similar. Therefore, using aerial imagery to distinguish between them can be harder. An example of this can be found in Figure 5, where distinguishing between the grassland and scrub, even manually, is difficult. On the other hand, woodland intensity characteristics are very distinguishable from the other habitats. Consequently, its recall ability is high, close to 65%, when retrieving the first 631 images.



**Figure 5.** (a) Grassland and (b) Scrub. Even though they belong to different habitat classes, their intensity properties are similar.

## 4.2 Classification

The classification accuracy of the method, shown in Table 2, was measured by applying k-NN and varying k, the number of neighbours taken into account when classifying the query image.

**Table 2.** Habitat classification using k-NN. Correctly classified images as k increases.

<i>Habitats</i>	<i>Values of k</i>												
	1	3	5	7	9	11	13	15	17	19	21	23	25
Arable	38	44	40	40	35	35	36	31	30	28	30	29	28
Grassland	163	122	23	16	16	15	17	17	18	15	15	14	15
Scrub	4	3	5	3	3	4	4	2	2	2	3	2	3
Woodland	68	123	140	157	164	169	167	171	172	177	182	182	183

As k increases, the number of correctly classified images decreases. This is particularly noticeable in grassland habitats where the classification accuracy drops from 122 with k=3 to 23 with k=5 as a consequence of intensity and characteristics similarities between different habitats, particularly scrub and grassland, previously commented in Section 4.1. On the other hand, results related to woodland habitats, whose characteristics are more distinguishable, increase as k increases, achieving a 70.5% of correctly classified photos when looking at the first 25 results.

## 5. Conclusions and further work

From the results showed in Section 4, it can be appreciated that aerial imagery and content-based image retrieval approach based on visual words and SIFT descriptors have its limitations in both retrieval and classification. The similarities between aerial images that represent different habitats, particularly grassland and scrub, present a problem when using visual words alone.

Further work includes the extraction of additional features, such as texture or information derived from the slope data. Moreover, instead of k-NN, which awards the same weight to all the results of the query regardless of their rank, a more refined computer vision algorithm for the classification of the habitats, such as random forest, could be implemented. Another alternative would be to evaluate the approach using multi-temporal images or a different set of photographs where habitat classes might be more distinguishable, e.g. ground-taken photography. This would take advantage from the fact that the codebook is independent from the test images.

Consequently, this approach can be seen as a starting point that, combined with further work such as other types of features or information (e.g. texture or slope) or the use of other types of computer vision methods in the decision making process (such as random forest), can be used to create an accurate automatic habitat classification algorithm.

## 6. Acknowledgements

All data courtesy of The Ordnance Survey. Particular thanks are due to Mr. Glen Hart, Dr. Carolina Sánchez-Hernández and Dr. Guoping Qiu for their help, guidance and encouragement. Mercedes Torres is supported by the Horizon Doctoral Training Centre at the University of Nottingham (RCUK Grant No.EP/G037574/1). This work was funded by Ordnance Survey and continues with funding from the RCUK's Horizon Digital Economy Research Hub grant, EP/G065802/1.

## 7. References

- Chen LC and Rau JY, (1997) Detection of shoreline changes for tideland areas using multi-temporal satellite images. *International Journal of Remote Sensing* 19(17), 3383-3397.
- Cover T and Hart P (1967) Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13(1), 21-27.
- JNCC (2010) Handbook for Phase 1 habitat survey - a technique for environmental audit, ISBN 0 86139 636 7.
- Qiu G (2002) Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition*, 35, 1675–1686.
- Sivic J and Zisserman A (2003) Video Google: A Text Retrieval Approach to Object Matching in Videos. *Ninth IEEE International Conference on Computer Vision (ICCV'03)*, Volume 2, 1470-1477
- Yang Y, Newsam S (2010) Bag-of-visual-words and spatial extensions for land-use classification. *SIGSPATIAL International Conference on Advances in Geographic Information Systems (2010)*. 270-279.

## 8. Biography

*Mercedes Torres studied Computer Science in the University of Seville and holds a MSc. in “Digital Signal and Image Processing” by Cranfield University. Currently, she is doing her PhD. in “Image Processing Applied to Habitat Classification” as part of the Horizon Doctoral Training Centre in the University of Nottingham.*