

Classification of Google Earth Entities

An investigation into how data mining techniques can be applied to satellite image data in order to identify, group and classify features within satellite imagery

Frans Coenen¹, Rob Christley²

¹Department of Computer Science The University of Liverpool, Liverpool, L69 3BX.

²Department of Veterinary Clinical Sciences, The University of Liverpool, L69 3BX.

26 June 2009

1. Background and Research Overview

Satellite image data is used in many application domains. Much of satellite imagery is publically available. However, one of the challenges of analysing satellite image data is the significant quantity of images that have been collected and are available for analysis. Much research work on satellite image analysis is therefore directed at the automated processing of such images. This current work is typically directed at specific applications: weather forecasting, deforestation monitoring etc. One category of application is the identification of features within satellite images, for example for security reasons or land usage monitoring. Whatever the case the analysis of satellite image features usually requires some form of segmentation and registration.

The proposed programme of work seeks, firstly, to apply image analysis techniques to satellite imagery to capture features within such imagery and secondly to apply data mining techniques to classify and cluster the identified features. Classification is concerned with the building of classifiers from labelled training data, which can then be used to classify new “unlabelled” data (this of course assumes the availability of appropriate training data). Classification is described as a supervised learning process, whereas clustering is described as an unsupervised learning process because it operates in the absence of training data. Clustering techniques seek to group data, features in this case, into “clusters”. When identified the definition of the clusters can be used to group further “examples”. The proposed work would seek to apply both classification and clustering techniques to the identified satellite image features.

More specifically the aim of the proposed research is to investigate how classification and clustering data mining techniques can be applied to analyse features with satellite imagery. Research issues include:

1. Mechanisms for processing satellite imagery to isolate features within those images and capture the details of those features.
2. The processing of the identified image features so that classification techniques can be applied.
3. The processing of the identified image features so that clustering techniques can be applied.
4. The presentation/visualisation of the results from 2 and 3.

2. Application Domain

To act as a focus for the research, and to provide a specific evaluation platform, the research will be directed at the analysis of “Google Earth” satellite imagery. More specifically the classification and clustering of “farm” features in rural parts of Ethiopia. This application has been chosen because the proposers have contacts “on the ground” in Ethiopia who can obtain labels for the desired training data. Part of the work would be to devise a procedure for the collection of such labels.

3. Rough Programme of work

A seven phase programme of work is envisaged. Each phase has deliverables associated with it that can be used as “milestones” to monitor progress.

Phase	Start (mnth.)	Duration (months)	Description	Deliverables
Previous work and background reading	1	4	Literature review of relevant topics and familiarisation with appropriate software systems	(1) Technical report(s)
Data capture	5	8	Investigation into data pre-processing	(1) Software, (2)

			procedure: segmentation and registration of images, feature identification, creation of training set (label capture)	Technical report, (3) Training set.
Clustering	13	6	Research into application of clustering techniques for feature grouping (unsupervised learning).	(1) Software, (2) Research Paper.
Classification	19	6	Research into application of classification techniques for feature categorisation (supervised learning).	(1) Software, (2) Research Paper.
Visualisation	25	3	Investigation of techniques with which to visualise the result	(1) Software, (2) Research paper
Final evaluation	28	1	Final evaluation of approach	(1) Research paper
Thesis preparation	29	8	Thesis writing and “plugging of gaps”.	(1) Thesis