CLASSIFICATION OF COCONUT FIELDS USING DATA MINING ON A LARGE DATABASE OF HIGH-RESOLUTION IKONOS IMAGES

V.P.Muthukumar[1], Dr.S.Subbaiah[2]

[1]Research Scholar& Assistant Professor,[2] Assistant Professor,

[1]PG & Research Department of Computer Science and Applications,

Vivekanandha College of Arts and Sciences for Women (Autonomous),

Tiruchengode, Tamilnadu, India.

[2]Department of Computer Science, Sri Krishna Arts and Science College,

Coimbatore, Tamilnadu, India.

ABSTRACT

Data Mining plays a vital role in research areas. Data mining is the process of extracting usable information from large amount of data. It is one of the most significant research areas in computer science. It is a calculation process of finding and determining valuable information from huge data set. Data Mining and Knowledge Discovery is a developing field of research that have been attracting many researchers to extract meaningful pieces of information from the dataset. In Remote Sensing Supervised classification of satellite images are commonly used. It is used to create the thematic maps based on a training set chosen by domain experts. These training set is called as ROI (Regions Of Interest) which is statistically characterize each class for e.g. coconut, sand of the satellite image. For each image, a set of ROI is manually created by domain expert. When we are using large number of images with high resolution will create difficulties of manual creation of ROI for each image. It is very time and money consuming process. In this paper, we proposed a method of semi-automatic approach based on clustering. It is to limit the number of ROI done by experts. After that, we use decision trees on a binary decomposition of RGB components for improving the classification. The Experiments have been done on 306 high resolution images of Tuamotu Archipelago.

INTRODUCTION

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems [1].

Our goal is to identify and estimate coconut fields in Tuamotu. Tuamotu archipelago contains 78 coralian atolls located be- tween 134 to 150 degree west and 14 to 24 degree south, covering 800,000 km^2 area. Based on the dimensions of the Tuamotu, coconut fields are of a great local economic importance for coprah and coconut oil production.

The segregation of those atolls and their dimensions make impossible a mission to obtain a complete inventory of all coconut fields. So, we plan to combine high resolution satellite images and supervised classification to deal with this problem. Our database consists in 306 high-resolution Ikonos Satellite Images

(0.80 m) covering 59 atolls (over 150 GB of data). These images were not taken at same times, and acquired with different climate and lighting conditions. They are composed of 3 RGB channels (Infrared is unfortunately lacking). Supervised classification is based on at least 10 classes such as vegetation, water (sea) and coral. However, extracting a number of ROI per image and per class is a tedious work considering the size of the database.

In this paper, we propose a semi-automatic approach based on data mining to identify coconut fields over these 306 high resolution images. This approach limits the amount of ROI to be done by experts for the classification. The proposed method is composed of the following steps:

- Creation of a set of ROIs per group of similar images. These groups are constructed using a clustering method on metadata (e.g. sun position, time, etc) and histograms.
- Supervised classification based on a decision tree technique where attribute-values of pixels are the binary decomposition of RGB components.

We validate our approach by comparing our results, for three atoll images, to the results of [2] which utilize image analysis on few Tuamotu atolls.

Section 2 presents some related works on tree segmentation method in satellite images. Section 3 gives the process of classification on huge database of high resolution images. Section 4 shows some experimental results on coconut classification on a huge database. Finally, section 5 concludes and gives some perspectives of the result.

II. RELATED WORKS

Several approaches exist for segmenting trees in satellite images. In [3] authors propose a method of finding local maxima of the Laplacian to detect the olive trees. Other meth- ods [4] try to extract trees in image that match synthetic pattern model. In [5], the authors use an heuristic method based on valleys detection (by applying masks in four direction). They consider image as a mountainous relief where the tree edges are local minimum. Similarly, [6] construct a network of edge points (local minimum calculated in four directions of 7x7 neighborhood) that correspond to tree crowns. Others researchers [7], [8] use the classical watershed method to segment trees. Perrin et al. [9] propose to extract the tree crowns by modeling forest stands by a process marked by ellipses.

A previous study was done by Raimana Teina in its thesis [9], this study is based on a part of the same Ikonos database that we use in this paper (mainly on the Tikehau atoll) but with a different goal. His work aims to detect coconut tree crown using watershed segmentation technics in order to estimate the number of coconuts in the whole Tuamotu archipelago. The segmentation results were compared to samples from a field campaign. This study showed that as not all coconut trees are visible on the Ikonos images, a detection rate has to be applied to quantify the number of coconuts really present on the ground, but this detection rate depends on the type of coconut field. Then, he introduced a coconut field typology and applied different classification algorithms in order to map precisely the different types of coconut fields as well as the other classes present on the images (sand, wet coral, dry coral, low vegetation, etc.). To achieve the separation of the different type of coconut fields, a set of classical texture indexes are used along with the

RGB bands.

The SVM (Support Vector Machine) classification method shows the best results among the other classical classification algorithms mainly because of the use of texture parameters along with RGB bands (which are not handled easily by classical classification algorithms such as Maximum Likelihood). The main drawback of the use of SVM algorithm is the very long computation times (several days for some images) which are not applicable to the entire Tuamotu Ikonos database. As our goal in this paper is to classify the whole Tuamotu Ikonos database using data mining algorithms, we use a simple class set (only one class for the coconut fields) and thus avoid the texture parameters computation which would inevitably increase the calculation time and add constraint on the classification algorithm. The Raimana Teina classification results on Fangatau, Nukutavake, Vahitahi, Vairaatea and Tikehau will be used as a comparison with our results to assess the quality of our method.

III. CLASSIFICATION OF A LARGE DATABASE OF HIGH RESOLUTION IMAGES

We describe in this section our approach is to classify a large database of high resolution images using data mining techniques (clustering and decision trees techniques). Fig.1 illustrates this process.

A. *Reducing the number of ROI creations*

In a first step, we identified all different objects (regions) generally appearing in atoll images (see fig. 2 and table.I). In a classification approach, given an image, experts would have to create an ROI for each different region. Thus

Number of ROI to construct by experts is $nbImg \times nbObj$,

with *nbImg is* the number of images to study and *nbObj is* the number of objects to classify. To avoid such a large number of ROI creations, the idea is to group similar images and to create only one set of ROI for this group.

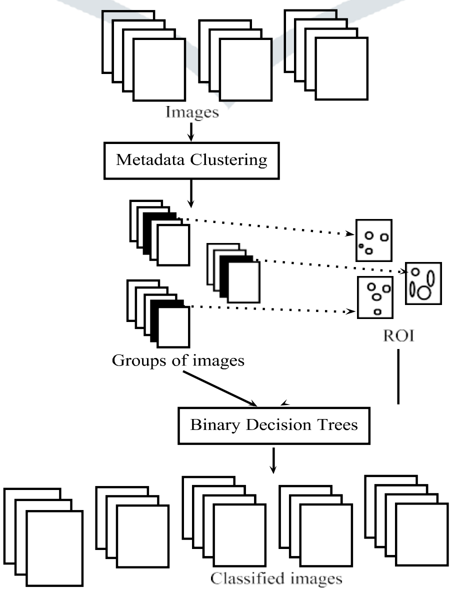


Fig. 1. Classification of a huge image database using clustering and decision trees

For example, if the group is composed of 20 images, experts would have to do only *nbObj* ROI (since we have *nbObj* classes of objects), instead of *nbImg* × *nbObj*. The question is how to group images?

Fig.2. Example of object classes

Images have dissimilar characteristics depending on acquisition time and position of the satellite. This affects the color and classification of these images. Generally, satellite images are provided with a set of metadata describing them and which may have a control on classification (see table II). We focus on properties that potentially impact the spectral response of objects. For example, we select information about the sun i.e. the azimuth and the elevation which have an effect on the lighting of the objects, and thus on their reflectance. The ocean surface may shine brightly according to the relative position of the sun with respect to satellite.

Table I: Classes Used

S.No	**Class**	**Color**
1	Coconut Field	Green
2	Other Vegetation	Dark Green
3	Deep Water	Blue
4	Swallow Water	Cyan
5	Wet Coral	Red
6	Dry Coral	Purple
7	White Sand	White
8	Waves	Yellow
9	Man Made	Orange
10	Cloud	Gray

Table II: Example of MetaData

Sun Angle Azimuth
Sun Angle Elevation
Nomiunal Collection Azimuth
Nominal Collection Elevation
Scan Azhimuth
Scan Direction
Acquired Nominal GSD

The weather (windiness for example) would have been interesting information to describe the data but this information is lacking. As a consequence, we studied other parameters that could affect significantly images. Among them, we focus on parameters derived from image histogram. The idea is to group images with the same dynamic and contrast. More precisely, we extract from the histogram the first and third quartiles, and the inter-quartile as new image metadata.

Based on these metadata, we create an instance per image with the seven parameters presented (i.e. acquisition time, position of the satellite, azimuth, elevation, first quartile, third quartile and inter-quartile). Then, a clustering method is used to group images w.r.t. these metadata parameters. The method used to construct the clusters is *X-means* algorithm [10] (implemented in Weka tool API [12]). This method is an improvement of the classical method *K-means*. In *K-means*, the number of clusters is fixed by user, while in *X-means*, it is automatically calculated in an optimal way.

For each cluster (i.e. group of images), we compute the centroid value w.r.t. seven metadata and histogram parameters. The image which is the closest of the centroid is chosen as the representative of the group. This representative is used by expert to define ROI. Classification of any image of the group will use these ROI extracted from the “representative image”. However all the images don’t necessarily have all the classes of the class set. For example, cloud, coconut plantation or other classes can be absent and induce classification errors. In such a case, we choose as “representative image” the second, the third, etc. closest image of the centroid.

B. Improving decision tree classification by using RGB binary decomposition

Decision trees allow prediction of an individual membership in a class according to its characteristics. A decision tree is a classification method equivalent to a set of decision rules. C4.5 is one of the major reference in decision tree classification [13], [14]. In our approach, we use J48 implementation (in the Weka Platform [12]) which is an optimized version of C4.5 algorithm.

We have, for classification, a set of objects corresponding to pixels of images. The previous extracted ROI are used as a basic training set. We first construct a tree using the three RGB components, then we consider the binary codes of these components. This decomposition will allow the improvement of our results. Indeed, each component (R, G and B) is encoded on 11 bits, we thus obtain 33 binary attributes. For example, a pixel with values “255 0 27” would be encoded ”0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1”.

IV. EXPERIMENTATIONS

A. *Experimentation protocol*

Unfortunatly we still have to plan a real ground truth mission in order to estimate the real performances of our method. Meanwhile, as the accuracy of R. Teina have been estimated and validated, in order to valid our approach, we compare our results, on four atolls (described in table III), with those obtained by R. Teina [9] with the same classes (i.e. the classes presented in table I). Results are compared w.r.t. coconut surfaces found. The idea is to evaluate quality of both approaches.

Table III: DESCRIPTION OF THE DATA USED IN OUR EXPERIMENTATIONS

ATOLL NAME	FANGATAU	NUKUTAVAKE	VAHITAHI	VAIRAATAE
pixel resolution	0.8m	0.8m	0.8m	0.8m
size in MBytes	916.115	1 409.824	942.453	598.426

We conduct the following two tests:

1) First, we study the quality of our binary classification method. In these experiments, we use a dedicated set of ROI for each image.

2) Second, we evaluate the impact of our approach to reduce the number of ROI on classification quality. In these experiments, we consider one set of ROI for a group of ”similar” image w.r.t. metadata. The impact of selected metadata on classification quality is also studied.

B. *Classification results with binaries attributes*

Table IV shows coconut surfaces found by our approach in comparison with the surfaces found in [2], [10]. Classification of images with the classes presented in table IV is visually satisfactory and surfaces of coconut plantations are closed to results obtained by [2], [10]. As we can see in table IV results are better with binaries values. Only image of Vahitahi atoll has an excessive coconuts surface. These classification errors are due to clouds, constructions, coral color and effect of wind and waves on the sea.

The coral classification is particularly problematic, because some coral area are labeled as ”coconut plantations”. We can see it in figure 4, which is a part of Vairaatae image, and where all green pixels are in reality coral. Coral may take many colors if it is dry or wet, including green. To correct this, we propose to replace the two classes ”wet coral” and ”dry coral” with ”green coral”, ”yellow coral”, ”grey coral”, and ”dark coral”. This leads to a better classification with less false coconuts. Indeed, coconut plantation surface of Vahitahi reduces to 3960642, 5 and the classification is improved.

Table IV: COCONUT PLANTATION SURFACES RESULTS (IN SQUARE METERS)

Fangatau	Nukutavake	Vahitahi	Vairaatae
6 070 097	3 403 047.5	3 711 583.2	1 954 993.6
4 877 639.5	3 355 070.5	4 197 018.5	2 005 360.9
4 809 770	3 315 760	2 579 720	2 439 110
+1.4%	+1.1%	+62.69%	-17.78%

As clouds are transparent, their spectral response changes significantly depending on floor below. Thus, good ROI are very difficult to realize and cloud are classified in many different classes including coconut plantations, as we can see on figure 3. The shadow of the cloud can lead to a false classification since pixels are darker. Modification of ROI would not be sufficient to correct this error; a solution is to preprocess images with classical methods of cloud processing.

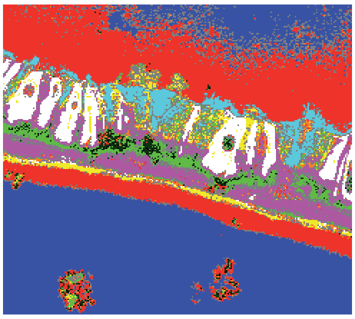


Fig. 3. Problem of cloud classification

"Man-made" constructions can take many different colors including some close to coral which can false other class's classification. To solve this problem, we classify images using the 11 classes presented in table I minus the "man-made" class. area. Note that "man-made" class is not really significantly extended in Tuamotu atolls. Those pixels are now labeled in diverse classes but not in the "coconut plantation" class, which is the only one of interest for our application.

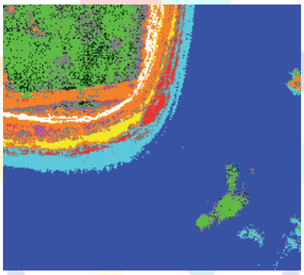


Fig. 4. Problem of coral classification

C. Results with a set of ROI per cluster

In this section, we group images by clustering on metadata and histogram parameters. Then, we use the most representative image of each cluster to create ROI, and use these ROI as input for the classification in 11 classes.

First, we test three combinations of metadata parameters for the clustering:

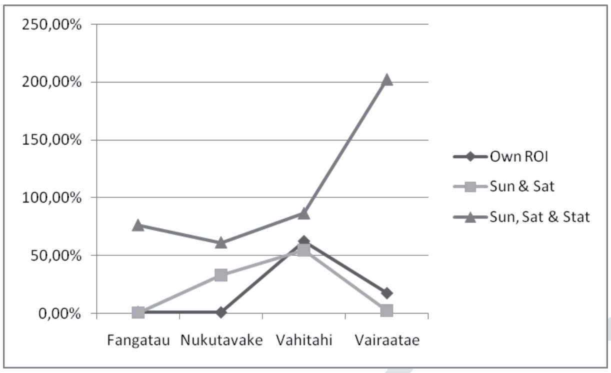
- sun parameters
- sun and satellite parameters
- sun, satellite and histogram parameters

Table V shows the impact of using ROI from metadata clustering on the classification. Note that the same classi- fication method is used for all these tests. We can easily see an important loss of quality compared to classification with images' ROI. It leads to an increase of classification's fuzziness and add glaring classification's errors.

Table V: PERCENTAGE OF DEVIATION FROM A CLASSIFICATION WITH ORIGINAL ROI

Sun	+0.39%	+127.64%	+488.73%	+38.44%
Sun and Satellite	-0.64%	-33.77%	-85.89%	+24.67%
Sun, Satellite and quartile	+73.93%	+59.40%	-42.34%	+267.82%

As shown by the table, the sun parameters greatly influence the image color, but are not sufficient enough to characterize images. Satellite parameters tends to improve results despite



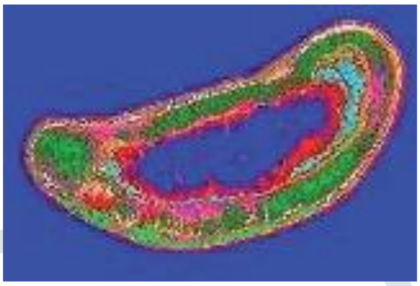


Fig.5. Percentage deviation from the surface

a big classification error on the atoll of Nukutavake which doesn't concern coconut plantation (see figure 6). The image quartiles and inter-quartile decrease the classification quality, in particular for coconut plantation surfaces. Other image statistics should be tested to improve it.

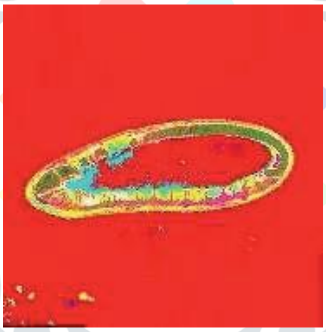


Fig. 6. Classification of Vahitahi with sun and sat parameters

As we have seen, our method results in lower classification quality than the classical "one set of ROI by image" classi- fications but the use of sun and satellite metadatas leads to promising results (see table V, figure 5 and figure 7). Time saving is really important with respect to classification with ROI of the image. Some improvement, as new statistics or new clustering method could improve the results.

V. CONCLUSION AND PERSPECTIVES

In this paper, we use the data mining methods to classify satellite images of the database. The first important point is the processing of a large database of images with a reduction of human involvement. Clustering of images based on their metadata creates image subdivisions sharing the same properties. Then, construction of Region Of Interests on only one representative image for each subgroup reduces the number of ROIs creation. This initial work shows hopeful results and leads us to investigate more in order to improve the classification enactments. For that, we recommend to investigate a supervised classification method based on the well-known noisy tolerant patterns [15]. A ground truth mission will have to be showed in a few Tuamotu atolls in order to assess more precisely the performances of our system.

VI. REFERENCES

[1] https://en.wikipedia.org/wiki/Data_mining# cite_note-acm-1

[2] R. Teina, D. Be´re´ziat, B. Stoll, and S. Chabrier, "Toward a global tuamotu archipelago coconut trees sensing using high resolution optical data," in *Proc. IEEE International Geoscience & Remote Sensing Symposium IGARSS'08)*, Boston, USA, Jan. 2008, pp. 290–294.

[3] K. Karantzalos and D. Argialas, "Towards automatic olive tree extraction from satellite imagery," in *Proc. of the XXth ISPRS Congress*, Istanbul, Turkey, July 12-23 2004.

[4] O. Kenneth, W. Jrgen, H. Johan, and O. Hakan, "Tree species discrim- ination using z/i dmc imagery and template matching of single trees," *Scandinavian journal of forest research*, vol. 21, no. 7, pp. 106–110, 2006.

[5] G. F. A. and L. D. G., "The individual tree crown approach applied to ikonos images of a coniferous plantation area," *Photogrammetric engineering and remote sensing*, vol. 72, no. 11, pp. 1287–1297, 2006.

[6] S. Culvenor, "Tida: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery," *Comput. Geosci.*, vol. 28, pp. 33–44, February 2002.

[7] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algo- rithm based on immersion simulations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 6, pp. 583 – 598, jun 1991.

[8] L. Wang, P. Gong, and G. S. Biging, "Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery," *Pho- togrammetric Engineering Remote Sensing*, vol. 70, no. 3, 2004.

[9] G. Perrin, X. Descombes, and J. Zeubia, "Evaluation des ressources forestie`res a` l'aide de processus ponctuels marque´s," in *Proc. Reconnais- sance des Formes et Intelligence Artificielle (RFIA06)*, Tours, France, 2006.

[10] R. Teina, "La caracte´risation de la cocoteraie des tuamotu a` partir d'images satellites a tree`s haute re´solution spatiale," Ph.D. dissertation, Univ. of Paris France, September 2009. [Online]. Available: http://hal.archives-ouvertes.fr/tel-00516952/

[11] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proceedings of the Seventeenth International Conference on Machine Learning*, ser. ICML '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 727–734. [Online]. Available: http://portal.acm.org/citation.cfm?id=645529.657808

[12] I. H. Witten and E. Frank, Data Mining : Practical machine learning tools and techniques (2nd edition). San Francisco, USA: Morgan Kaufmann Publishers Inc., 2005.

[13] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.

[14] Programs for Machine Learning. Morgan Kaufmann, 1993

[15] D. Gay, N. Selmaoui-Folcher, and J.-F. Boulicaut, "Application independent feature construction based on almost-closedness properties," Knowledge and information Systems, pp. 1–25, 2010, 10.1007/s10115-010-0369-x. [Online]. Available: http://dx.doi.org/10.1007/s10115-010-0369-x53