Proceedings of initial ECO-IMAGINE GI and GIS for Integrated Coastal Management, Seville 13th–15th May 2004

Automatic delineation of shoreline and lake boundaries from Landsat satellite images

Stefano Bagli and Pierre Soille
Joint Research Centre of the European Commission
Institute for Environment and Sustainability - Land Management Unit
via E. Fermi 1, Ispra, Italy
{Stefano.Bagli,Pierre.Soille}@jrc.it

Abstract

Motivated by the need to generate a pan-European coastline and water bodies database from Landsat 7 ETM+ images, we present a new methodology for extracting automatically the coastline and the lake boundaries. Our approach consists of the combination of spectral and spatial information for the images using morphological image segmentation techniques. For these purposes several morphological segmentation algorithms were implemented inside a GIS platform to evaluate their performance in coastline and lake boundaries extraction. Preliminary results for Ireland and UK demonstrate the accuracy of the developed methodology as well as it's applicability to a large areas.

1. Introduction

The delineation and extraction of coastlines and water bodies, e.g. rivers and lakes, from remote sensing image is an important task useful for various application fields such as coastline erosion monitoring, coastal zone management, watershed definition, GIS database updating, flood prediction, and the evaluation of water resources. Tracing the coastline manually, although easy along relatively simple stretches of coast, is not practical where the coastline becomes very complex. In addition, automatic and replicable techniques are required to update coastline maps, to evaluate the spatial and temporal evolution of alterations due to natural and anthropic events, and to extract the waterline for large areas.

The extraction of features, such as coastlines and water bodies directly from satellite images overcomes the problem of matching available coastline data sets with the studied image data set. In fact, owing to projection system biases, the matching of a water boundary coming from a different data set together with the available images may turn out to be a tedious if not impossible task. A scheme to detect shoreline changes using manual digitisation of multi-temporal satellite images with tidal measurements and DEM was presented by Chen [3]. Beyond manual digitisation, several techniques have been reported in the literature for the derivation of the coastline position from satellite images [7], [2], [4], and [6]. They usually rely on density slice using single or multiple bands and multi-spectral classification, both supervised and unsupervised (e.g. ISODATA, PCA, Tasseled Cap, NDWI). These approaches are mostly based on evaluation of the spectral signature in the near and middle infrared portion of the electromagnetic spectrum. In this range the water bodies absorb almost all incident radiant flux while the land surface reflects significant amount of near and middle infrared energy.

Following the approach proposed by Wilson [9] and motivated by the need to generate a pan-European coastline and lakes database for environmental applications at the JRC, we present a new methodology for automatically extracting the coastline and lakes boundaries. Our approach consists of the combination of spectral and spatial information for the images using morphological image segmentation techniques.

2. Methodology

The proposed methodology has two major components. The first is the preprocessing and mosaicing procedure of extracted images pertaining to a specific region, e.g. nation or watershed. The preprocessing step regards also the extraction of clouds and their shadows from the images in order to minimise the presence of clouds and their shadows in the mosaic. The second involves extracting waterline by using morphological image segmentation algorithms. A flowchart of the proposed scheme is shown in figure 1.

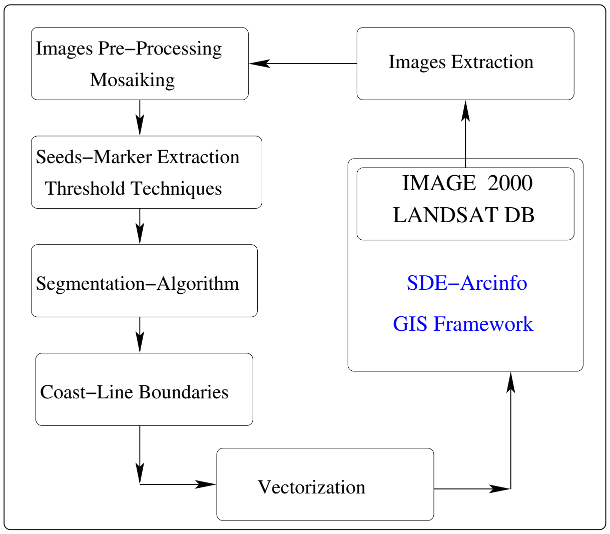

Figure 1. Flow chart of the methodology

2.1. Preprocessing

Pan-European Landsat satellite data were collected by EC-JRC inside a project called Image2000, in order to update the European Corine Land Cover database.

Image2000 images were produced from ETM+ Landsat 7 satellite, except few images taken from Landsat 5, providing both multi-spectral and panchromatic data. The images were delivered already orthorectified in national projection systems, then resampled using the cubic convolution and collected during the reference year 2000 with a deviation of maximum 1-year.

The preprocessing procedures consist of extracting the Landsat images pertaining to the region of interest, e.g. nation state or regions, using the ArcSDE Raster Engine (ESRI) from the JRC Image Oracle DB and mosaicing them in a specific projection system. The developed mosaic algorithm permits the removal of clouds and areas of shadows from overlapping image regions, if one of the available images is cloud free. The detection of clouds and their shadows is based on both spectral and spatial criteria.

2.2. Segmentation

The algorithm for segmenting an image into meaningful regions requires some prior knowledge about the image objects that are to be recognised. We tested two different morphological segmentation algorithm. The first is a modified version of the Seeded Region Growing (SRG) algorithm initially proposed by Adams and Bishof [1] and later enhanced by Mehnert and Jackway [5] to achieve independence on the order in which pixels are processed. SRG is based on the postulate of region growing algorithms, where a criteria of similarity of pixels is applied. Instead of controlling region growing by tuning homogeneity parameters, SRG is controlled by choosing a small number of pixels, called seeds. It starts with the selection (manual or automatic) of a number of seed regions $R_1, R_2, \ldots, R_n$ to which a region growing technique is applied. At each step, the algorithm proceeds by adding one unassigned pixel to one of the above sets until there are no more unassigned pixels in the image. This algorithm can be formally explained as follows. Let $N(\mathbf{x})$ denote the set of immediate neighbours of the pixel $\mathbf{x}$, and T the set of yet unallocated pixels which border at least one of the regions. That is, T corresponds to the external morphological gradient ρ^+ (see e.g. Soille [8]) of the union of all R_i:

$$T = \rho_B^+(R). \tag{1}$$

where $R = \bigcup_{i=1}^n R_i$ and the neighbour set B is defined as a 3×3 square structuring element. If, for any pixel $\mathbf{x} \in T$, the dissimilarity d between $\mathbf{x}$ and R_i can be defined as follows:

$$d(\mathbf{x}, R_i) = \begin{cases} |f(\mathbf{x}) - \mathrm{mean}_{y\in R_i}\{f(y)\}|, & \text{if } \mathbf{x} \in \rho_B^+(R_i), \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

where $f(\mathbf{x})$ is the grey level value of the pixel $\mathbf{x}$. Since a given pixel of T may border more than one region, we define the dissimilarity between this pixel and the set of all regions as the minimum dissimilarity between this pixel and all regions:

$$d(\mathbf{x}, R) = \min_i d(\mathbf{x}, R_i). \tag{3}$$

In each iteration of SRG algorithm we assign the pixels of T which have the minimum dissimilarity to their corresponding region. That is, all pixels $\mathbf{z}$ of T such that $d(\mathbf{z}, R) = \min_{\mathbf{x}\in T}\{d(\mathbf{x}, R)\}$ are appended to their corresponding R_i. We proceed until all image pixels have been assigned to a region. The algorithm produces a tessellation of the image into the same number of regions, as those given by the seed regions. The boundaries of each homogeneous region are extracted by determining the boundary of the corresponding region using a morphological gradient function. In order to better evaluate the performance of SRG algorithm in waterline extraction from satellite images, we applied different combination of seed regions criteria (using multiple band) with different SRG algorithm versions.

The second algorithm is based on the watershed transformation (WS) using a marker-controlled segmentation procedure. The watershed transformation provides a clustering around the image minima.

Both SRG and WS segmentation algorithms require the identification of seed or marker regions. Rather than letting the user defining these regions, complete automation is achieved using a simple method of extraction for seed and marker regions, such as threshold techniques, and using the result of the first segmentation as a starting point for SRG or WS. In our case, each seed and marker regions belongs to

Table 1. Threshold values used in seed and marker regions for land and water.

Slice Density	Sea/Water	Land/Soil
Band5	1-12	101-255
Pan	1-20	70-255
Band7	1-12	81-255
NDWI35	0.4-0.47	-0.25 - -0.1

either sea/water or land/soil. These regions were identified automatically on multispectral and panchromatic Landsat images using a simple density slice (threshold techniques) on single or multiple (NDWI) bands [4], [2]. The threshold values for extracting land and water seeds were derived empirically on test images and they remain valid for the entire area of interest. Table 1 shows a list of threshold values used for automatic seed and marker region extraction for different bands and NDWI. The main requirement for identifying seeds and markers, consists of the delineation of representative pixels for each class without including noisy pixels which can invalidate the seeded region growth and the marker segmentation procedure.

2.3. Result Evaluation

An index of performance was developed in order to identify the best performing waterline extraction algorithm. This quantitative index evaluates the success of the segmentation methodology in delineating coastline and lakes water boundary with respect to an existing reference layer, coming from available geographic data sets. The performance index (PI) is obtained by computing the total area of discrepancies between reference and automatically extracted coastline or lake.

The PI is defined as in equation 4:

$$\mathrm{PI} = 1 - (\sum_{i}^{n} A_i/\mathrm{Npix}) \tag{4}$$

where A_i is the generic area of discrepancies and Npix correspond to the total number of pixel contained in a buffer computed using a fixed distance from the reference layer. The performance index is evaluated in a test area with complex coastline and lakes shape located on the west coast of Norther Ireland (NIE). The reference coastline and lakes features for this area is a layer from MOLAND (Monitoring Land Use / Cover Dynamics), a research project carried out at the IES - JRC. The Moland coastline and lakes boundaries for NIE were extracted by a manually digitising high resolution (5 m), IRS satellite images. Another data set,

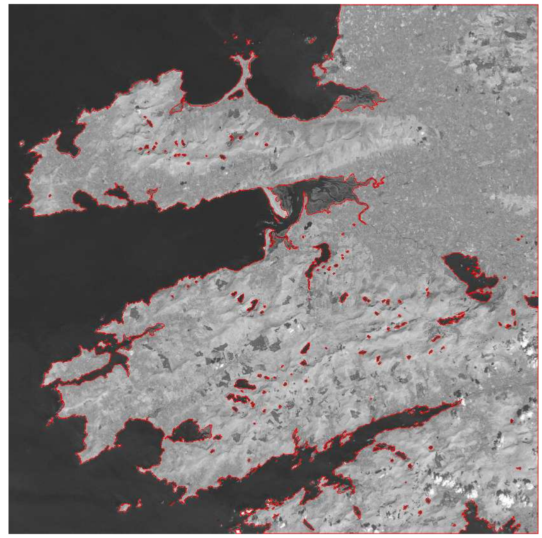

Figure 2. SRG (red) from Landsat 7 ETM+ Band 5 Coastline

representing a reference coastline and lake layer for the European Commission, stored in a GISCO database was used for PI computation. Table 2 displays the results of the performance of different segmentation algorithm computed using the Moland coastline/lake layer, the last row of the table show the PI between GISCO and Moland coastline/lake.

Analysing the PI computation it appears that the both SRG and WS algorithm on band 5 with seed regions extracted from the same band delineate, land/sea and lake boundaries, with good match with respect to the Moland reference Coastline. The performance index PI represents just an indicator to evaluate the performance of different segmentation algorithms, but it cannot be considered as an absolute indicator for the coastline extraction accuracy. In fact it refers to a manually digitised reference coastline that can be subject to human error. In order to evaluate correctly the absolute accuracy of extracted coastline reference land/sea border points from ground surveys (such as GPS observation with simultaneous Landsat 7 ETM+ observation) [6] would need to be available.

Some results of lake boundaries and shoreline extraction are displayed in figures 3 and 2.

From visual analysis of the above maps, it appears that the waterline delineated using morphological operators fits very well the actual coast and lake detected from the Landsat Satellite images.

3. Conclusion

The results of this work demonstrate that region growing and morphological segmentation can be applied to Landsat

Table 2. Shoreline and Lake Boundaries Performance Index

Segmentation Method	Seeds or Marker	Segmentation	Coastline PI	Lake PI
SRG	DS band 5	band 5	97.43	99.30
SRG CORE	DS band 5	band 5	97.40	99.29
SRG GRAD	DS band 5	band 5	96.43	99.26
SRG	DS band 5	band 5-3-7	96.86	/
SRG	DS NDWI band 3-5	band 5	96.67	/
WS	DS band 5	band 5	97.33	99.33
GISCO DB	/	/	97.06	97.87

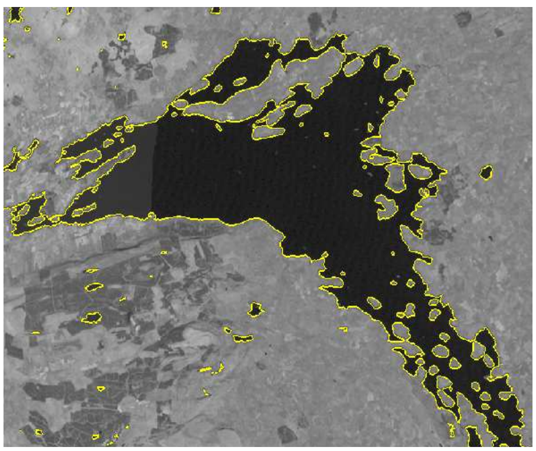

Figure 3. WS (yellow) from Landsat 7 ETM+ Band 5 Lake boundaries

7 ETM+ data in order to delineate and map coastline for the entire pan-European Continent. The methodology developed is completely automatic and produces vector files of the coastline which can be analysed and mapped using GIS tools for European Commission needs. The evaluation of the results through a computation of a performance index demonstrates that the coastline extracted automatically is better that available data sets, such as GISCO coastline.

4 Acknowledgements

The authors are grateful to A. de Jager for his help in mosaicing and database application, to V. Lima for providing Image2000 Landsat images, C. Lavalle for providing Moland coastline and all the LMU IT staff.

References

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transaction on pattern analysis and machine intelligence*, 16(6):641–647, June 1994.

[2] D. H. Braud and W. Feng. Semi-automated construction of the Luisiana coastline digital land/water boundary using Landsat Thematic Mapper satellite imagery. Technical Report 97-002, Department of Geography & Anthropology, Luisina State University. Luisiana Applied Oil Spill Research and Development Program, OSRAPD, 1998.

[3] L. C. Chen and J. Y. Rau. Detection of shoreline changes for tideland areas using multi-temporal satellite images. *ijrs*, 19(17):3383–3397, 1998.

[4] P. S. Frazier and K. J. Page. Water body detection and delineation with landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66(12):1467–1467, Dec. 2000.

[5] A. Mehnert and P. Jackway. An improved seeded region growing. *Pattern Recognition Letters*, 18:1065–1071, 1997.

[6] J. Ryu, J. Won, and K. D. Min. Waterline extraction from Landsat TM data in a tidal flat. A case study in Gomso bay, Korea. *Technical Report Series Department of Geography & Antropology Lousiana State University*, (83):442–456, 2002.

[7] L. Smith. Satellite remote sensing of river inundation area, stage, and discharge: a review. *Hydrological Processes*, (11):1409–1413, 1997.

[8] P. Soille. *Morphological Image Analysis*. Springer-Verlag, Heidelberg, 2nd edition, 2003.

[9] P. A. Wilson. Rule-based classification of water in Landsat MSS images using the variance filter. *Photogrammetric Engineering and Remote Sensing*, 63(5):485–491, 1997.