Vectorization and spatial query architecture on island satellite images
Suleyman Eken a *, Ahmet Sayar a

a Kocaeli University, Department of Computer Engineering, TR-41380 Umuttepe, Kocaeli, Turkey

Abstract

Vector data structure produces smaller file size than raster image because a raster image needs space for all pixels while only point coordinates are stored in vector representation. Raster-to-vector conversion, i.e. vectorization, has key functionalities in Geographic Information Systems (GIS). This study presents a technique converting raster satellite island images into vector data as polygons by using edge-detection algorithms. This enables storing satellite images as vector objects into spatial database systems and utilizing the advanced built-in spatial functions and queries developed by the spatial database community. The effectiveness of the technique is demonstrated by applying on a system detecting spatiotemporal changes in a real-world island satellite image.

Keywords: Vectorization, Raster Data, Vector Data, Spatial Databases, Spatial Queries, Data Modelling;

1. Introduction

There are two kinds of computer graphics – raster and vector. Raster graphics are also called bitmap images, and consist of a grid of individual pixels where each pixel can be a different color or shade. On the other hand, vector graphics use mathematical relationships between points and the path connects them to describe an image. Although each representation has advantages over another, they can be used together in many applications [1]. Raster-to-vector conversion, i.e. vectorization, has key functionalities in Geographic Information Systems (GIS), image processing for Remote Sensing, conversion of technical drawings (analog to digital) in engineering applications, and document analysis/recognition.
Inasmuch as applications are generally internet based, it is necessary to transfer data from one node to another. Depending on data sizes and limited network bandwidth, distributed applications might not give the expected results in terms of response times and scalability. When it comes to applications using images as data sets, then the issues become more serious. To produce a raster image file accurately, graphics software keeps track of a large amount of information, including the exact location and color of each pixel in the collection of pixels. This results in huge file sizes for raster graphics. Spatial queries such as area, length, circumference, proximity, and neighbourhood require vector information more than raster information. Performing such queries on vector images result in low computation processes and high performance. Using vector representations of raster images solves some of those problems and enables high performance computations in distributed information system applications in which images are analyzed with spatial queries.

In this paper, vectorization is realized through the use of edge detector. Method is based on applying Canny operator to greyscale image to find out the edges, and then, converting the edges into binary-level with Otsu threshold technique. In the next step, it apply 4-way scan to find out the border pixels of the image, and create a convex hull covering those pixels by aid of Gift Wrapping algorithm. The technique is applied on an island satellite image. This enables storing satellite images as vector objects into spatial database systems and utilizing the advanced built-in spatial functions and queries developed by the spatial database community.

The remainder of this paper is organized as follows. In Section 2, relevant works are presented. In Section 3, vectorization based on edging is proposed. Section 4 gives some experimental results and Section 5 draws a conclusion.

2. Relevant Works

Many vectorization methods have been developed and implemented. These methods are shortly classified in six groups: Hough Transform based methods parameterize straight lines according to its slope and intercept [2], skeletonization based methods [3], contour based methods [4], graph-structure based methods [5], mesh pattern based methods [6], and pixel tracking based methods [7]. With the exception of the Hough Transform based methods a typical vectorization process consists of medial axis representation acquisition, chain coding, and line segment approximation or polygonization.

Medial axis points sampling is the core processing for information reduction by means of eliminating trivial points which do not represent the core-line. With line tracking process, one pixel wide skeleton points are used to generate poly-line. Polygonization part includes that some intermediate points on the straight line segment formed by their neighbours are removed from the chain list while maintaining the original shape [8].

Hough transformation based methods are very time consuming as every edge pixel influences every other pixel. Many papers have been written on improving the speed and accuracy of the Hough transformation [9-11]. Major disadvantages of most thinning algorithms are high time complexities, loss of shape information (such as line width), distortions at junctions, false and spurious branches, and sensitive to noise [12]. Contour based algorithm is robust and capable of dealing with all line shapes without misjudging a junction and producing an incorrect skeleton. Hence, it is inappropriate for use in vectorization of curved and multi-crossing lines [13]. Drawbacks of run graph based methods are incorrect intersection points and undesirable junction areas [8]. Because of this reasons they are
not suitable for curve line vectorization. Owing to fact that controlling mesh size is crucial, mesh based methods are not appropriate for dashed lines. The pixel tracking based methods are noise sensitive and does not perform well vectorization of arc images because they vectorize raster curve image to a set of bars.

Most of the vectorization algorithms are applied on straight line images [14, 15]. As we deal with island images which are not composed of only straight lines like engineering drawings, we need a novel method. Architecture of our system is explained at Section 3.

3. Architecture

We first apply Canny operator to greyscale image to find out the edges, and then, convert the edges into binary-level with Otsu threshold technique. In the next step, we apply 4-way scan to find out the border pixels of the image, and create a convex hull covering those pixels by aid of Gift Wrapping algorithm. In the last step, points of the convex-hull are stored into the database as a polygon and corresponding image is made ready for the spatial queries. All steps are shown in Figure 1.

![Fig. 1. General Approach](image)

3.1. Vectorization of Raster Island Images

Raw island image is converted into grayscale for simplifying the tasks of the subsequent steps in the algorithm (Step 1 in Figure 1 and its output is Figure 2.b). Applying Canny edge detection algorithm [16] to grayscale image may significantly reduce the amount of data to be processed and may therefore filter out information that may be regarded as less relevant, while preserving the important structural properties of the image (Step 2 in Figure 1 and its output is Figure 2.c).

The success of a vectorization process is directly proportional to the quality of a raster image. The quality of an image mostly depends on how clearly image represents the real world object. The resolution and some outer factors such as noises (overcast sky, dark sky) also affect clarity of an image. Since these are defined as outer obstacles and cannot be controlled in the system, we try to use noiseless images to get best results. Additionally, algorithm specific parameters such as thresholding value and order of directions in 4-way scan step affect the achievement of vectorization. Otsu thresholding technique [17], taking gray level values distributions in image as well as the local characteristics of the pixels into consideration, was used to separate The Island of Cyprus from the rest of the image (Step 3 in Figure 1 and its output is Figure 2.d).

After Otsu thresholding step, each pixel in the image including part of The Island of Cyprus were coded as 1(white pixel), others as 0 (black pixel). When the image is scanned from left to right line by
line horizontally, the 1s encountered first will be left-hand-side border pixels of The Island of Cyprus image. When the image is scanned from right to left line by line horizontally, the 1s encountered first will be right border values of the Island image. When the image is scanned from top to down line by line vertically, the 1s encountered first will be upper-hand-side border pixels of the Island image. When the image is scanned from bottom to up line by line vertically, the 1s encountered first will be lower border pixels of the Island image. (Step 4 in Figure 1 and its output is Figure 2.e).

We can represent islands by polygons. Polygons can be represented mathematically by a circular sequence of points providing that starting and ending points are of the same (x, y) values. Even if the edge detection and four-way scan algorithms determine the points representing the island, points must be ordered in order to create a polygon.

Convex hull corresponds to the intuitive notion of a boundary of a set of points and can be used to approximate the shape of complex object. In other words, convex hull of a set of points S is the boundary of the smallest convex region that contains all the points S inside it or on its boundary. To find polygon, covering all points in the plane with small area, representing the island, Gift Wrapping algorithm was used [18]. After this stage, raster island image was converted to vector island image (Step 5 in Figure 1 and its output is Figure 2.f).

3.2. Storing Vector Island Images into Spatial Databases

Spatial databases store and query data for spatial objects in accordance with object relational database model. While typical databases can only understand various numeric and character types of data, spatial databases can analyze spatial data types such as point, line, and polygon. Spatial data can be geo-spatial maps, rivers, roads, and islands that can be represented with points, lines easily or satellite images, digital elevation models, and aerial photographs formed groups of pixels. Typical SQL queries such as how many population Kocaeli has, how many towns Kocaeli has totally, and which districts beginning with the letter A of Kocaeli are can be query on spatial databases. In addition, typical queries, spatial queries such as where two closest restaurants to engineering faculty (proximity) are, which town intercity bus terminal in Kocaeli (containment) is, what Turkey’s neighboring countries (adjacency) are, railway through which towns of Kocaeli passes (intersection-overlap) can be analyze on spatial databases.

For the proof of correctness and analysis of the proposed technique we utilize PostgreSQL and PostGIS. PostGIS is an extension to the PostgreSQL object-relational database system which allows GIS objects to be stored and queried in the databases. The Island of Cyprus image in vector form is recorded PostgreSQL database with non geometric features such as island’s name, area, and length and geometric feature such as polygon representing the island.
We handle a polygon having five-vertex; to store this polygon into PostgreSQL spatial database we can use these SQL expressions:

CREATE TABLE polygon_region (name varchar, coordinate geometry);
INSERT INTO polygon_region VALUES ('island', 'POLYGON((0 10,15 10,15 -5,10 -5,10 0,0 10))');

After storing island images as polygons into spatial database, we can realize such kind of spatial operators shown in Table 1 on geometry class:

<table>
<thead>
<tr>
<th>Operators</th>
<th>Operator Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Envelope</td>
<td>Returns the minimum bounding rectangle</td>
</tr>
<tr>
<td>Closed</td>
<td>Returns true if the LINESTRING's start and end points are coincident</td>
</tr>
<tr>
<td>Boundary</td>
<td>Returns the boundary of the geometry</td>
</tr>
<tr>
<td>Intersection</td>
<td>Returns true if geometries are not disjoint</td>
</tr>
<tr>
<td>Contains</td>
<td>Tests whether the given geometry contains another given geometry.</td>
</tr>
<tr>
<td>Overlap</td>
<td>Returns true if the interiors of two geometries have non-empty intersections</td>
</tr>
<tr>
<td>Area</td>
<td>Returns the area of the surface if it is a polygon or multi-polygon</td>
</tr>
<tr>
<td>Length</td>
<td>Returns the length of the geometry if it is a line string or multilinestr</td>
</tr>
</tbody>
</table>

4. Case Study and Evaluation

Some scenarios such as earthquake (Scenario-1), erosion (Scenario-2) and filling (Scenario-3) are applied on a satellite image for an island and their effects are calculated via proposed techniques. The effects of these phenomena named as scenarios are measured by the values of circumferences and area before and after the simulations. The area and circumference values are obtained by spatial queries on vector representation of the image. Image is stored as a polygon in an object relational PostgreSQL database.

The real area and length (or circumferences) are calculated from the polygon coordinates according to the below formula:

Real_length = Map scale value * Length calculated from polygon in database
Real_area = (Map scale value)^2 * Area calculated from polygon in database

Sample spatial queries performing area and length of vector island image are listed below:

SELECT Round(ST_Area(formerpolygon)), Round(ST_Area(latterpolygon)), area FROM dbpoligon WHERE ad= 'Cyprus'
SELECT Round(ST_length(ST_Boundary(formerpolygon))), Round(ST_length(ST_Boundary(latterpolygon))), contour FROM dbpoligon WHERE ad='Cyprus'

In scenario-1, we obtained artificial earthquake with translocation operation on The Island of Cyprus image. We studied on how much slip in km has occurred by comparing centroids of polygons obtained from images before and after translocation. (See the arrow in Figure 3.a)

In scenario-2, artificial erosion has been applied on the island by deleting a portion of The Island of Cyprus image. Then, we have investigated how much reduction has occurred by querying the area and the length of the polygon. (See the circle in Figure 3.b)
In scenario-3, artificial filling was obtained by adding a portion to the Island of Cyprus. Then, we investigated how much expansion has occurred by querying the area and the length on polygons. (See the circle in Figure 3.c)

Table 2. Result of simulations

<table>
<thead>
<tr>
<th>SCENARIOS</th>
<th>Area (km²)</th>
<th>Length (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Slipping</td>
<td>15715</td>
<td>15715</td>
</tr>
<tr>
<td>Erosion</td>
<td>15715</td>
<td>14016</td>
</tr>
<tr>
<td>Filling</td>
<td>15715</td>
<td>17380</td>
</tr>
</tbody>
</table>

The actual area of The Island of Cyprus is 9251 km² and its length is 648 km. The area of the convex representation of the island image is 15715 km² and its length (actually circumference) is 543 km. Convex representation of the island image include some parts which are not existed in the actual island image. This is only an approximation of the real world island, and even if it does not give exact results for the queries such as area and length it gives some hints about the changes happens in its geographic coordinates. By considering this fact, when we look at the values in Table 2 we see that there is no change in the area and circumference values of the island image before and after the application of scenario-1 (slipping). In this scenario, degree of slip was determined as 25 km by comparing the centroids of the polygon representing the image before and after the translocation. In scenario-2 (erosion), it is shown that the area and circumference of the image has decreased around 1699 km² and 8 km respectively. When it comes to scenario-3 (filling), it is seen in the table that area of the image has expanded 1665 km² and a length of the image has increased 14 km. In addition, there is no slipping in scenario-2 and scenario-3 due to the lack of translocation.

5. Results and Future Works

The study presented in this paper proposed a technique to convert a raster image into a vector data as an object by using edge detection algorithms. This enables storing satellite island images into spatial database systems as polygons and utilizing built-in spatial tools and functions developed by spatial database community. Application is made on a real satellite image and effectiveness of the proposed technique is examined with simulations. In this study we have tested spatial queries such as how much slip has occurred, how much reduction has occurred, how much expansion has occurred on The Island of Cyprus image after the applications of the scenarios via a model based on Canny edge detection technique and obtained successful results.
With the world of Internet and development of related technologies, applications of satellite imagery and other raster images in web based distributed systems have increased rapidly. Such applications with limited network bandwidth and internet infrastructure do not allow transfer and processing of high sized images effectively. As the recommended approach reduce image size extremely it will contribute to service-oriented distributed applications.

Concave representations of the island images give much better results in terms of spatial queries such as area and length. However, when the images have second or more degree curved structure (i.e., non-reachable from 4-way directions, north-south-east-west), it is very hard to create a concave representation. In the future, we will improve the algorithm to solve this problem. Also, resolution and some outer factors such as noises (overcast sky, dark sky) also affect clarity of an image. We will try to use noisy images in the future.

References