University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications from the Department of
Electrical and Computer Engineering

Electrical & Computer Engineering, Department
of

2003

A Shape-Based Approach to Change Detection of Lakes Using Time Series Remote Sensing Images

Jiang Li

University of Nebraska-Lincoln

Ram M. Narayanan

University of Nebraska-Lincoln

Follow this and additional works at: <https://digitalcommons.unl.edu/electricalengineeringfacpub>

Part of the [Electrical and Computer Engineering Commons](#)

Li, Jiang and Narayanan, Ram M., "A Shape-Based Approach to Change Detection of Lakes Using Time Series Remote Sensing Images" (2003). *Faculty Publications from the Department of Electrical and Computer Engineering*. 149.

<https://digitalcommons.unl.edu/electricalengineeringfacpub/149>

This Article is brought to you for free and open access by the Electrical & Computer Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Department of Electrical and Computer Engineering by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A Shape-Based Approach to Change Detection of Lakes Using Time Series Remote Sensing Images

Jiang Li, *Member, IEEE*, and Ram M. Narayanan, *Fellow, IEEE*

Abstract—Shape analysis has not been considered in remote sensing as extensively as in other pattern recognition applications. However, shapes such as those of geometric patterns in agriculture and irregular boundaries of lakes can be extracted from the remotely sensed imagery even at relatively coarse spatial resolutions. This paper presents a procedure for efficiently retrieving and representing shapes of interesting features in remotely sensed imagery using supervised classification, object recognition, parametric contour tracing, and proposed piecewise linear polygonal approximation techniques. In addition, shape similarity can be measured by means of a computationally efficient metric. The study was conducted on a time series of radiometric and geometric rectified Landsat Multispectral Scanner (MSS) images and Thematic Mapper (TM) images, covering the scenes containing lakes in the Nebraska Sand Hills region. The results validate the effectiveness of the proposed processing chain in change detection of lake shapes and show that shape similarity is an important parameter in quantitatively measuring the spatial variations of objects.

Index Terms—Change detection, polygonal approximation, remote sensing, shape analysis.

I. INTRODUCTION

SHAPE RETRIEVAL and representation in remote sensing imagery have not received adequate attention as in other fields such as computer vision [1]. It is relatively easy to recognize the regular shapes such as farm fields that often appear as squares or rectangles, or highways and roads that are either straight or curved lines. However, irregular shapes such as lake boundaries tend to be scribbles and polygons without a set pattern. Recently, the point diffusion technique (PDT) [2], proposed to handle objects whose shape can be represented by a set of sparse points, was applied to meteorological satellite images for fast and efficient shape similarity evaluation. Comprehensive overviews of the state of the art in shape analysis are given in [3] and [4], which discuss techniques such as image segmentation, object recognition, and contour extraction, as well as algorithmic solutions, e.g., approximation of shapes and shape similarity measure.

Shape pattern analysis of water-covered surfaces in the Western Lakes Region of the Nebraska Sand Hills is potentially useful for flood and drought monitoring. Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) images provide researchers with unique chronological records for studying surface water resources where no other hydrologic data exist.

Buckwalter's work [5] stands as the original effort to monitor variations in the surface-water area of Sand Hills lakes. MSS data have been used to study the interrelationships between precipitation events and changes in the surface-water area of certain lakes [6] and to investigate the seasonal and interannual patterns of lake area variability [7]. No study has addressed lake shape change, an important parameter in characterizing its spatial variations. Furthermore, the segmentation procedure for delineating surface water involves only simple level thresholding of intensity in the near-infrared band [8]. The challenge is to determine what specific level of the band reflectivity marks the "threshold" between terrestrial cover and standing surface water. This motivates us to seek a better segmentation of water bodies using the information in full bands of the multispectral images.

Techniques proposed for grayscale image segmentation in computer vision are generally edge-based or region-based. However, it is inherently difficult to extend differential methods to multiband remotely sensed images by computing the gradient of the vector field to obtain edges of regions [9]. A hierarchical segmentation algorithm [10] using region growing and spectral clustering and a hybrid approach [11] to integration of edge and region data in image segmentation exist. Unfortunately, the practical implementations of these algorithms need large-scale parallel processing to achieve reasonable processing times. In this study, we expect that classification over the multispectral data will achieve a satisfactory segmentation. Furthermore, the *a priori* knowledge available about the study region, e.g., the location of each lake of interest, allows us to identify appropriate training sites and use a supervised classification approach based on support vector machines (SVM) rather than an unsupervised classification technique, e.g., clustering or mathematical morphology [12]–[14]. The SVM is a novel type of learning machine based on statistics learning theory introduced by Vapnik [15] and applied by many others [16]. SVM has shown superior performance in real-world applications [17] including classification of remote sensing imagery [18]–[20].

Water body masks need to be created from the classified images to obtain a series of binary images containing only objects of interest, e.g., lakes, which can be identified by the connected components. Tanimoto [21] employed a recursive algorithm that works on one component at a time. Dellepiane and Fontana [22] proposed a fuzzy approach to intensity connectedness. Other algorithms were designed for large images that may not fit in memory [23], [24]. An adapted version of the algorithm presented in [25] was used in this study.

The approaches to shape representation can be broadly divided into transform-based, region-based, and contour-based

Manuscript received August 20, 2002; revised May 13, 2003. This work was supported by the Nebraska Research Initiative Geospatial Data Analysis project.

The authors are with the Department of Electrical Engineering, University of Nebraska, Lincoln, NE 68588-0511 USA (e-mail: ram@doppler.unl.edu).

Digital Object Identifier 10.1109/TGRS.2003.817267

Fig. 1. Block diagram of the proposed processing chain.

[4]. Transform-based methods such as Fourier coefficients are usually used as shape descriptors for classification, while representing shapes directly by regions requires large storage space. Therefore, we adopted the parametric contour-based approach that represents the shape outline as a parametric curve implying a sequential order along it.

To quantify the shape change, we resorted to an efficiently computable shape similarity metric [26], which requires parametric contours be approximated as polygons. The ideal procedure is to represent the contours by polygons with minimum number of vertices under a certain fit criterion [27]–[29]. However, it remains an open problem to find an algorithm that returns the fewest number of vertices, yet runs in $O(n)$ time. In this study, we developed a polygonal approximation algorithm running in $O(n)$ time and giving a small but nonminimum number of vertices, adapted from the piecewise linear interpolation algorithm [30], [31] for time series data.

This paper introduces a processing chain to efficiently retrieve and represent shapes of interesting objects, e.g., lakes, in remotely sensed imagery, and to detect changes in those shapes. Section II gives an overview of the processing chain. Section III discusses the SVM-based supervised classification. Information on connected components labeling and parametric contour tracing is presented in Section IV. Sections V and VI describe the proposed piecewise linear polygonal approximation algorithm and the shape similarity metric, respectively. Section VII discusses our results, while Section VIII concludes with proposals for future work.

II. OVERVIEW OF THE PROCESSING CHAIN

Fig. 1 shows the block diagram of the proposed processing chain. We employed 36 four-band MSS images (256×256 pixels) from 1981 to 1987 and 10 six-band TM images (768×768 pixels, bands 1–5, 7) from 1992 to 1997, covering scenes containing the lakes of the Nebraska Sand Hills region. All the images were precalibrated and registered to a UTM-13 map.

We classified each image using the developed optimized SVM classifier. Then we applied connected component labeling on the binary images of water body masks created

TABLE I
REGION-GROWING ALGORITHM

```

repeat until  $A_R > A$ 
  for each pixel  $p$  at the border of  $R$  do
    for all neighbors  $p_n$  at  $(r_n, c_n)$  of  $p$  do
      if  $\|(r_n, c_n) - (r_s, c_s)\| \leq D$  and  $\|\mathbf{I}(r_n, c_n) - \boldsymbol{\mu}_R\| \leq S$  then
        add  $p_n$  to  $R$  and update  $\boldsymbol{\mu}_R$ 
      end if
    end for
  end for

```

from the classified images. After obtaining the contours and corresponding features (e.g., centroid and perimeter of each lake of interest) from the retrieved components via parametric contour tracing, we converted the contours into polygons using the proposed piecewise polygonal approximation approach. Finally, we applied the shape similarity measure to quantify changes in the shapes.

III. SUPERVISED IMAGE CLASSIFICATION

We developed a region growing algorithm to sample the training and test data according to spatial constraints (i.e., maximum area A of the region and maximum distance D from the seed) and spectral constraints (i.e., spectral Euclidean distance S , referring to the distance between the spectral vector of a specific pixel and the mean vector of pixels in the region). Other similarity criteria incorporate neighboring pixels to regions such as spectral angle distance (SAD), which is invariant to unknown multiplicative scalings of spectra that may arise due to differences in illumination and angular orientation [12]. However, these effects do not seem important in the images used in this study considering the classification results, although they might be relevant in other applications. The elements of the spectral vector, composed of the gray level in each of the four bands $[I_1, I_2, I_3, I_4]$ for MSS data and in each of the six bands $[I_1, I_2, I_3, I_4, I_5, I_7]$ for TM data, are normalized to $[0, 1]$. The procedure is outlined in Table I, where (r_s, c_s) is the coordinate of the seed, $\mathbf{I}(r_n, c_n)$ refers to the spectral vector at a pixel (r_n, c_n) in region R , and $\boldsymbol{\mu}_R$ is the mean vector of all pixels in R .

As a general rule [32], $10n$ pixels of training data for each class need to be sampled where n is the number of bands. We sampled more than 100 pixels for each class in an MSS image and 200 pixels in a TM image. After interactively selecting seeds via a friendly graphical user interface (GUI), we fixed A and adjusted D and S until enough data were sampled. For example, we chose $A \geq 100$ with $50 \leq D \leq 100$ for an MSS image. The region is more sensitive to the parameter S . A large S incorporates pixels of other classes while a small S causes the algorithm to stop quickly with few data sampled. Therefore, we need to decide whether the regions should be merged if multiple detections are present for a single region due to a strict similarity criterion. Current implementation of the GUI-based Signature Editor allows us to easily merge or remove a region from the sampled datasets. Five datasets for each class were sampled at different sites, and we used 20% of the data for training and the remaining 80% for testing.

Fig. 2. (a) Original Landsat TM image. (b) Classified Landsat TM image.

A. Structure of the SVM Classifier

A comprehensive description of SVM is given in [17], which is summarized as follows. Given a set of examples consisting of pairs of class labels and n -dimensional feature vectors as $(y_i, \mathbf{x}_i), i = 1, \dots, l, y_i \in \{1, -1\}, \mathbf{x}_i \in \mathbf{R}^n$, the SVM approach places the hyperplane $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0, \mathbf{x} \in \mathbf{R}^n, b \in \mathbf{R}$, so that the margin, defined as the distance of the closest vectors in both classes to the hyperplane, is maximized. The hyperplane is obtained by the optimization problem

$$\min_{\mathbf{w}, b} \langle \mathbf{w} \cdot \mathbf{w} \rangle \quad y_i (\langle \mathbf{w}_i \cdot \mathbf{x}_i \rangle + b) \geq 1, \quad i = 1, \dots, l \quad (1)$$

which can be translated into the following form by introducing the Lagrange multipliers $\alpha_i \geq 0$:

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \quad (2)$$

subject to $\sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, l$.

Only a small number of multipliers α_i have nonzero values and they are associated with the so-called support vectors which form the boundaries of the classes. This maximal margin classifier can be generalized to nonlinearly separable data via two approaches. One is the introduction of a soft margin parameter C to relax the constraint of all the training vectors of a certain class lying on the same side of the optimal hyperplane. This approach is effective in case of noisy data. The other is the transformation of input vectors into a higher dimensional feature space by a map function φ , followed by a linear separation there. The expensive computation of inner products can be significantly reduced by using a suitable kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$.

We implemented the SVM classifier using the SVM LIB library [33] and adopted radial basis function (RBF) defined as the kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2). \quad (3)$$

A practical difficulty of using SVM is the selection of parameters, e.g., the soft margin parameter C and kernel parameter γ in our case. Although SVM is not sensitive to different choices

TABLE II
LAND COVER CATEGORIES

Land Cover Type	Color
Open Water	
Emergent Wetlands	
Bare Rock/Sand/Clay	
Grasslands/Herbaceous	
Pasture/Hay	

of parameter γ for RBF kernel [34], it is desirable to obtain optimal values for both of these parameters given a special dataset. Automatic model selection for SVM has been studied extensively in the field of machine learning. Several upper bounds of the generalization errors were defined [35], [36], and efficient algorithms [37], [38] were designed to search the best values for these parameters. Accuracy in this study is satisfactory using $C = 100$ and $\gamma = 0.5$, the optimal values computed by LOOMS (leave-one-out model selection) algorithm [39]. According to USGS Land Use/Land Cover map, five major land cover classes are identified in Table II. Fig. 2(a) and (b) shows an original TM image and the corresponding classified image respectively.

B. Accuracy Assessment

The classification was repeated five times using a different 20% subset of the sampled data for training and the remaining 80% for testing. Table III gives an average performance for different classes for an MSS image and a TM image. Note that the producer's and user's accuracy are both high and water is almost perfectly classified. In general, results obtained from TM images are better than those from MSS images. Table IV shows that the overall classification accuracy of support vector machine classifier outperforms those of the maximum-likelihood classifier and the minimum-distance-to-means classifier.

IV. OBJECT RECOGNITION AND CONTOUR TRACING

After creating a water body mask from the classified image to obtain a binary image containing objects of interest [lakes in

TABLE III
CLASSIFICATION ACCURACY FOR DIFFERENT CLASSES

Class Name	MSS Image		TM Image	
	Producer's Accuracy	User's Accuracy	Producer's Accuracy	User's Accuracy
Open Water	98.4%	98.7%	99.8%	99.7%
Emergent Wetlands	91.7%	92.6%	93.3%	93.6%
Bare Rock/Sand/Clay	89.5%	87.8%	90.4%	88.1%
Grasslands/Herbaceous	88.3%	86.4%	89.0%	86.8%
Pasture/Hay	92.1%	94.5%	94.9%	96.4%

TABLE IV
OVERALL PERFORMANCES FOR DIFFERENT CLASSIFIERS

Classifier	Overall Classification Accuracy	
	MSS Images	TM Images
Support Vector Machine	91.5%	93.6%
Maximum Likelihood	87.2%	89.8%
Minimum Distance to Means	80.6%	82.3%

Fig. 3(a)], we applied a labeling algorithm [25] to identify all the different sets of pixels that are connected to each other. The algorithm scans through a binary image and locates connected regions defined by “1” pixels against a background of “0” pixels. Once finishing labeling all the “1” pixels reachable from the first found “1”, it then creates another label and searches for the next unlabeled “1” pixel, repeating until all the “1” pixels in the image are exhausted.

The area of each object is calculated by counting the number of pixels of each connected components during the labeling. The histogram of the average area (Fig. 4) is used to assist the selection of nontrivial objects. Objects with average area size larger than a threshold, e.g., 200 pixels for MSS images and 1000 pixels for TM images, are considered for further representation. In particular, the smallest rectangle parallel to the coordinate axis that encloses a nontrivial object (the so-called Feret Box) is computed during labeling as shown in Fig. 3(b), wherein the name of each studied lake is indicated.

Existing contour tracing algorithms such as run-length codes [40] and chain-codes [41] assume that only one object lies in a given image. To handle multiple objects represented by the connected components in each binary image, we developed an algorithm that automatically examines all the Feret boxes associated with the objects to extract their contours and features such as centroids and perimeters. Our algorithm is based on an efficient contour following algorithm [4] that can trace out the external contour of a single connected object in a binary image as briefly described as follows.

An initial point on an object’s contour is selected first by locating an unlabeled pixel that has a labeled right neighbor using line-by-line raster scanning. Fig. 5(a) shows an example wherein the pixel at position S denotes the starting point surrounded by eight neighbors, of which the pixel at position 0 is a labeled pixel. We assume that positions 1, 2, and 3 are removed from initial consideration as the next contour point as these have been verified not to be contour points during the scanning process. Fig. 5(b) shows the remaining four possible candidates for the next contour point at positions 4, 5, 6, and 7. At

each step, the algorithm remembers the direction from the current contour point to the next contour point and tests the neighbors in counterclockwise order. This process is repeated until the contour tracing is completed when the initial point is revisited. The pseudocode of this algorithm is presented in [4], which is not further discussed here.

Since the Feret boxes might overlap, the tracing algorithm regards only pixels with the current component label as the object and all other pixels as background. The location of each lake can be automatically recognized by referring to its UTM coordinates [7]. Specifically, we converted the UTM coordinates of each lake to relative coordinates within a binary image by mapping the UTM coordinate of the pixel at the upper left corner to the origin. The centroid of the lake is computed as the average complex value of all its points, $c[n] = x[n] + iy[n]$, where $x[n]$ and $y[n]$ are their contour coordinates. The names of the lakes are assigned to the Feret boxes with minimum Euclidean distance between centroids and reference coordinates. Afterwards, each lake is characterized by perimeter, shape (contour), and area. The complete algorithm is summarized in Table V. Fig. 6(a)–(c) shows the variations of the shape of the contours retrieved for Crescent Lake on different dates over a two-year period.

V. PIECEWISE LINEAR POLYGONAL APPROXIMATION

The contour is converted into a polygonal before a shape similarity metric is applied to quantitatively measure the changes of a lake boundary. In addition, polygonal approximation eliminates noise (spur pixels) introduced by the classification errors and reduces the total data that needs to be stored. We developed a piecewise linear polygonal approximation (PLPA) algorithm running in $O(n)$ time, adapted from the piecewise linear interpolation algorithm for time series data presented in [30] and [31]. This approximation algorithm produces a small but non-minimum number of vertices that lie on the given contour presented by an array of points.

Given a sequence of points $S : (x_0, y_0), \dots, (x_{N-1}, y_{N-1})$ and an error threshold Ψ , PLPA finds a piecewise linear function f , composed of a set of segments represented by linear functions $f_j, j = 1, \dots, K$, where K is the total number of pieces whose domains are disjoint. Assuming that the start and end points of a piece are $[x_m, y_m]$ and $[x_n, y_n]$ respectively, as shown in Fig. 7(a), each linear function f_j is defined as

$$f_j(x) = \frac{y_n - y_m}{x_n - x_m}(x - x_m) + y_m. \quad (4)$$

Fig. 3. (a) Binary image containing water. (b) Lakes of interest with Feret boxes.

Fig. 4. Histogram of the average area of each lake.

The lower line and upper line according to the threshold Ψ are denoted as

$$L_j(x) = \frac{(y_n - \Psi) - y_m}{x_n - x_m}(x - x_m) + y_m \quad (5)$$

$$U_j(x) = \frac{(y_n + \Psi) - y_m}{x_n - x_m}(x - x_m) + y_m. \quad (6)$$

PLPA tries to extend each piece until the upper line U_j becomes lower than the lower line L_j . Fig. 7(b) illustrates the case $x_n < x_m$, in which L_j starts from (x_m, y_m) and passes through $(x_n, y_n + \Psi)$ while U_j starts from the same point and passes through $(x_n, y_n - \Psi)$. Table VI summarizes the algorithm. The input is a point sequence S (the coordinates of the contour) and a maximum error threshold Ψ , and the output is a piecewise linear polygonal approximation function. The variable *Start* is the start point with S_L and S_U representing the slope of a piece.

The approximated polygons of the contour of Crescent Lake in June 1985 under different thresholds are shown in Fig. 8(a)–(c). It can be seen that the smaller the Ψ , the more segments the piecewise linear function (i.e., the polygon) will contain, thus preserving more details of the shape.

The compression ratio (CR) defined for different thresholds as

$$CR = 1 - \frac{\text{number of polygon vertices}}{\text{number of contour points}} \quad (7)$$

is shown in Fig. 9(a). In this case, the larger the threshold value, the greater the number of points along the contour that can be eliminated.

The threshold theoretically measures the maximal distance between the original points and the approximated line. Suppose the length of the perimeter is P , and the upper bound of the approximation error in the area is $e_A \leq P \cdot \Psi$. Using data from different representative lakes, Fig. 9(b) shows the approximation precision (AP) defined as

$$AP = 1 - \left| \frac{A - \tilde{A}}{A} \right| \quad (8)$$

where A represents the original area and $\tilde{A}$ is the approximated value for different compression ratios. Lakes with a high perimeter-to-area ratio (PAR), defined as P/A , show higher errors in general than those having low PAR. Note that very high precision is conserved even at a 55% compression ratio. Therefore, we chose 0.25 as the threshold.

VI. SHAPE SIMILARITY MEASURE

Assuming that the noise is roughly uniformly distributed over the sides of the approximated polygon, we can apply a shape similarity metric, a function defined on pairs of shapes indicating the degree of their resemblance, to give a quantitative measure of shape changes of a lake on different dates.

Veltkamp and Hagedoorn [3] review popular metrics, e.g., Hausdorff distance, turning function, and reflection metric for varied shape representations such as finite point sets, curves, and regions. Other important region-based metrics, such as oversegmentation and noise region, are discussed in [42]. In this study, we adopted an efficiently computable metric [26] based on the

Fig. 5. (a) Verified positions during scanning search. (b) Automatic contour tracing on connected components.

Fig. 6. Retrieved contours of Crescent Lake. (a) June 1983. (b) June 1984. (c) June 1985.

TABLE V
PARAMETRIC CONTOUR TRACING AND LAKE RECOGNITION

```

for each Feret box  $F_i$  do
  Trace and save the object's contour
  Compute centroid  $C_i$  and perimeter  $P_i$  of the object
  for each lake with reference coordinate  $RC_j$  do
     $d_{ij} = \text{Euclidean\_Distance}(C_i, RC_j)$ 
    Assign object the name of the lake with  $\min(d_{ij})$ 
  end for
end for

```

L_2 distance between the turning functions of the two polygons. The metric has several important properties. First, it is invariant under translation, rotation, and change-of-scale. Second, it can be computed in time $O(mn \log mn)$ where m is the number of vertices in one polygon, and n is the number of vertices in the other. Finally, it is insensitive to small perturbations and

matches human intuitive notions of shape resemblance. A detailed description of the metric is presented in [26] and is briefly described below.

Given a polygon A represented by a list of vertices around its boundary with corresponding coordinates, the turning function $\Theta_A(s)$ is the cumulative angle of the counterclockwise tangent as a function of the arc length s , measured from some reference point O on A 's boundary. It increases with left-hand turns and decreases with right-hand turns. The polygon should be rescaled so that $s \in [0, 1]$, and $\Theta_A(s)$ is a function from $[0, 1]$ to $\mathcal{R}$. Since $\Theta_A(s)$ may become arbitrarily large over $[0, 1]$ for a concave polygon, we should have $\Theta_A(1) = \Theta_A(0) + 2\pi$ in order to represent a closed polygon [26]. Therefore, given two polygons A and B and their associated turning functions $\Theta_A(s)$ and $\Theta_B(s)$, we can measure the degree to which they are similar by taking the distance between $\Theta_A(s)$ and $\Theta_B(s)$ in terms of the metric on function spaces. Suppose the turning function is $\Theta_A(s+t)$ if the reference point O along A 's boundary is shifted by t , and it

Fig. 7. Start and end points of a segment (a) $x_n \geq x_m$ (b) $x_n < x_m$.

TABLE VI
PIECEWISE LINEAR POLYGONAL APPROXIMATION ALGORITHM

```

Start =  $(x_1, y_1)$ ,  $S_L = -\infty$ ,  $S_U = +\infty$ 
for  $i = 1$  to  $n - 1$  do
     $S'_L = \max(S_L, \text{slope}(\text{Start}, (x_{i+1}, y_{i+1} - \Psi)))$ 
     $S'_U = \min(S_U, \text{slope}(\text{Start}, (x_{i+1}, y_{i+1} + \Psi)))$ 
    if  $x_{i+1} \leq x_i$  then
        Exchange( $S'_L, S'_U$ )
    end if
    if  $S'_L \leq S'_U$  then
         $S_L = S'_L$ ,  $S_U = S'_U$ 
    else
         $f_i(x) = \frac{S_L + S_U}{2}(x - \text{Start}.x) + \text{Start}.y$ 
         $\text{Start} = (x_i, f_i(x_i))$ 
         $S_L = \text{slope}(\text{Start}, (x_{i+1}, y_{i+1} - \Psi))$ 
         $S_U = \text{slope}(\text{Start}, (x_{i+1}, y_{i+1} + \Psi))$ 
        if  $x_{i+1} \leq x_i$  then
            Exchange( $S_L, S_U$ )
        end if
    end if
end for

```

is $\Theta_A(s) + \theta$ if A is rotated by θ , then to minimize the distance overall of such effects, the similarity measure is [26]

$$d_2(A, B) = \left(\min_{\theta \in R} \int_0^1 |\Theta_A(s + t) - \Theta_B(s) + \theta|^2 ds \right)^{\frac{1}{2}}. \quad (9)$$

VII. STUDY RESULTS

All of the algorithms discussed in this paper have been implemented with Visual C++ in Windows, and analyses have been conducted via a friendly interactive GUI. Experiments were carried out on selected lakes with three objectives: 1) to look at seasonal and annual variations in area and perimeter, 2) to seek similarities and differences in the spatial variability according to centroid and geometric shape, and 3) to determine the statistical relationship between the lake surface area changes and drought events.

A. Study Results Using the MSS Images

Using the MSS dataset, we examined the average variability of the surface area occurring in consecutive months for studied lakes in terms of relative values as shown in Fig. 10. Generally, the lakes experience surface area maxima in June with minima occurring in October, which is consistent with the pattern described in [6]. Note that on average, the area of Goose Lake shows a considerable increase during May and June and a significant decrease during August and September.

To examine the annual fluctuations, we calculated the coefficient of variation $CV = \sigma/\mu$ [7] for the area and the perimeter of each lake for each year, where μ is the mean and σ is the standard deviation. We then took an average of these coefficient values as shown in Fig. 11. Note that Island Lake and Goose Lake have $CV \geq 10\%$ for both the area and the perimeter, while other lakes show relatively small annual fluctuations. An interesting discovery is that CV of the perimeter is always less than that of the area except for Goose Lake, which might be caused by its more irregular shape with a higher PAR.

Spatial variability of each lake was inspected based on centroids and shape similarity measures on different dates. For example, Fig. 12(a) shows the centroid variations of Crescent Lake from month to month with corresponding shape similarity measures in the year 1983. For shape similarity measurement, a value less than 0.5 is a minor change, a value between 0.5 to 1.0 shows medium change, and a value larger than 1.0 reflects significant variation. Fig. 12(b) shows the interannual variations of Crescent Lake occurring in the same month (July) in different years. These quantitative measurements of the spatial variation allow us to easily identify the lake experiencing the most significant spatial change—useful hydrologic information unavailable using traditional analysis approaches. Our study results reveal that in general, Island Lake, Hackberry Lake, and Goose Lake's shapes show more variations than those of other lakes, and the shape of Goose Lake presents the largest variability. We also discovered that the variations of centroids are within only several pixels along both axes, with each pixel representing 79 m, the spatial resolution of MSS images.

A good correlation exists between lake surface area and precipitation occurring over the previous 45 days, and this pattern tends to be repeatable and steady [6]. We are interested in the relationship between the lake area changes and the drought events. We correlated area variations of each lake with the short-term (one-month) modified Palmer drought severity index (PDSI) [43], which has been used in the National Drought Atlas as a

Fig. 8. Approximated polygons of the contour of Crescent Lake in June 1985 under different thresholds (a) $\Psi = 1.00$. (b) $\Psi = 0.50$. (c) $\Psi = 0.25$.

Fig. 9. (a) Compression ratio for different thresholds. (b) Precision for different compression ratios.

Fig. 10. Average seasonal variability of lake area in 1983.

Fig. 11. Average coefficient of variation of lake area and perimeter.

good representation of existing conditions for real-time operational use. Fig. 13 illustrates the variation of the PDSI in different years. Drought is defined as beginning in the month when the PDSI equals or falls below -1 after having been above -1 . Drought duration is defined as the interval of time for which the PDSI remains equal or below -1 [43]. Fig. 14 shows the correlation coefficient between the areas and the indices for each lake. Strong correlations are observed although the coefficients

are not always positive. The figure shows positive correlation when the indices stay below -1 , i.e., within the drought period as in 1981 and 1987. Negative correlations are observed at the end of the drought cycle, i.e., the indices go above -1 as in 1983 and 1984. A very strong correlation occurred in 1985 when all the indices are above -1 , which indicates a normal period.

The reason for the significant diverse patterns of correlations in different years is not readily obvious to us. Lacking knowl-

Fig. 12. (a) Spatial variation of Crescent Lake in 1983. (b) Spatial variation of Crescent Lake in July.

Fig. 13. Variation of PDSI.

Fig. 14. Correlations between lake area and PDSI.

edge of potential casual factors and hydrologic parameters such as the complicated local topography, possible ground-water recharge or transmission area, and lake basin sediments in the study area [7], we are not able to provide further discussion of possible cause and effect relationships here.

B. Study Results Using the TM Images

We studied three more lakes: Wolf Lake, Bean Lake, and Swan Lake, taking advantage of higher spatial resolution of TM images. However, due to the small number of TM images available to us, results obtained from TM data only show changes occurring in summer (June versus August) and autumn (September versus October).

Figs. 15 and 16 illustrate that Crescent Lake, Island Lake, Hackberry Lake, and Blue Lake have no significant area change in summer except in 1993, when all the lakes except Wolf Lake show a primary area reduction. All the lakes show nearly the same area fluctuation pattern in autumn except Ashburger Lake, which shows a considerable gain, and Bean Lake, which shows a remarkable loss. These exceptions might be caused by the quick charge or discharge of underground water. With regard to perimeter, Goose Lake and Bean Lake show substantial variability while others have similar patterns in autumn. Swan demonstrates more fluctuations of both area and perimeter than others in the summer but not in autumn.

Fig. 17 shows that the pattern of the shape similarity in summer is more significant than that in autumn. Generally, Blue Lake and Hackberry Lake have their shape preserved as with a similarity measure around 0.5, although Hackberry Lake attains a value of 1.4 in the summer of 1995. Ashburger Lake and Wolf Lake show shape similarity variability from 0.7 in autumn to 1.0 in summer, thus experiencing moderate changes in shape. Note the variability of the shape of Island Lake in summer is larger than in autumn, while Goose Lake's shape changes considerably in both summer and autumn. Comparing Fig. 17 with Figs. 15 and 16 shows significant shape variation does not always imply large area or perimeter change and vice versa.

The variation of the centroid of each lake along both axes, on average, is within six pixels with each pixel representing 30 m, the spatial resolution of TM images. This fluctuation, similar to that obtained using MSS images, is not regarded as significant. Therefore, we conclude that the centroids of the studied lakes are basically reserved although the boundary shapes of some lakes experience somewhat strong variations.

It is important to point out the effects of the spatial resolution in this study. Overall, images with better spatial resolution allowed us to study some relatively small lakes and disclose more details about the boundary change of the lakes, which in turn increased our confidence about the corresponding detected spatial variations. Therefore, although multispectral images were the only datasets available to us for the study area, it is possible to apply similar shape analysis to high spatial resolution grayscale images using alternative segmentation approaches.

Fig. 15. (a) Surface-area fluctuations of the lakes in summer. (b) Perimeter variability of the lakes in summer.

Fig. 16. (a) Surface-area fluctuations of the lakes in autumn. (b) Perimeter variability of the lakes in autumn.

Fig. 17. (a) Shape similarity of the lakes in summer. (b) Shape similarity of the lakes in autumn.

VIII. CONCLUSION

We have presented a shape-based approach to lake change detection using time series multispectral images. Algorithms adopted in the processing chain have proved to be effective for

retrieval and representation of objects' shapes in the remote sensing images. It has been shown that the optimized support vector machine classifier offers performance improvements over traditional classifiers. The use of automated processing procedures such as connected components labeling and

parametric contour tracing has generated satisfactory results. An efficient piecewise polygonal approximation method has been developed and applied to approximate and represent the boundary of the lakes.

Study results have shown that shape similarity should be considered important in quantitative measurement of the spatial variations of objects, in addition to the classical parameters such as area and perimeter. However, the limitation of the temporal resolution of the available image datasets prevents us from performing a more thorough trend analysis. The correlation between the lake area and the drought indices is regarded as significant in specific cases. Future study should use images with higher spatial resolution to characterize the shapes of the objects more precisely. The proposed approach can be used to automatically retrieve and represent other types of objects with irregular shapes, though it has been applied to a special object of interest, the lake, in this study.

ACKNOWLEDGMENT

The authors thank P. Z. Revesz [University of Nebraska-Lincoln (UNL)] for his help in the development of the piecewise linear polygonal approximation algorithm and D. C. Rundquist (UNL) for providing the image datasets. The authors also thank J. Mitchell (Cornell University) for providing the source code for the shape similarity measure.

REFERENCES

- [1] J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, Berlin: Springer-Verlag, 1993, pp. 130–131.
- [2] F. Dell'Acqua and P. Gamba, "Query-by-shape in meteorological image archives using the point diffusion technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 1834–1843, Sept. 2001.
- [3] R. C. Veltkamp and M. Hagedoorn, "State-of-the-art in shape matching," Utrecht Univ., Utrecht, Netherlands, Tech. Rep. UU-CS-1999-27, 1999.
- [4] L. F. Costa and R. M. Cesar, *Shape Analysis and Classification: Theory and Practice*. Boca Raton, FL: CRC, 2001.
- [5] D. W. Buckwalter, "Monitoring Nebraska's Sandhills lakes," Conservation and Survey Div., Inst. Agriculture and Natural Resources, Univ. Nebraska-Lincoln, Lincoln, NE, Resource Rep. 10, 1983.
- [6] D. C. Rundquist, M. P. Lawson, L. P. Queen, and R. S. Cerveney, "The relationship between summer-season rainfall events and lake-surface area," *Water Resources Bull.*, vol. 23, no. 3, pp. 493–508, June 1987.
- [7] D. C. Gosselin, D. C. Rundquist, and S. K. McFeeters, "Remote monitoring of selected ground-water dominated lakes in the Nebraska Sand Hills," *J. Amer. Water Resource Assoc.*, vol. 36, no. 5, pp. 1039–1051, Oct. 2000.
- [8] E. A. Work and D. S. Gilmer, "Utilization of satellite data for inventorying prairie ponds and lakes," *Photogramm. Eng. Remote Sens.*, vol. 42, no. 5, pp. 685–694, 1976.
- [9] A. Cumani, "Edge detection in multispectral images," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 40–51, Jan. 1991.
- [10] J. C. Tilton and W. T. Lawrence, "Interactive analysis of hierarchical image segmentation," *Proc. IGARSS*, vol. 2, pp. 733–735, July 2000.
- [11] J. Le. Moigne and J. C. Tilton, "Refining image segmentation by integration of edge and region data," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 605–615, May 1995.
- [12] A. Weisberg, M. Najarian, B. Borowski, J. Lisowski, and B. Miller, "Spectral angle automatic cluster routine (SAALT): An unsupervised multispectral clustering algorithm," *Proc. IEEE Aerospace Conf.*, vol. 4, pp. 307–317, Mar. 6–13, 1999.
- [13] L.-K. Soh and C. Tsatsoulis, "Segmentation of satellite imagery of natural scenes using data mining," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1086–1099, Mar. 1999.
- [14] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 309–320, Feb. 2001.
- [15] C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 1–25, 1995.
- [16] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Disc.*, vol. 2, no. 2, pp. 121–167, 1998.
- [17] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [18] S. Fukuda and H. Hirose, "Support vector machine classification of land cover: Application to polarimetric SAR data," in *Proc. IGARSS*, vol. 1, Sydney, Australia, July 2001, pp. 187–189.
- [19] J. A. Gualtieri, S. R. Chettri, R. F. Crompt, and L. F. Johnson, "Support vector machine classifiers as applied to AVIRIS data," in *Summaries 8th JPL Airborne Earth Science Workshop*, R. Green, Ed., Pasadena, CA, Feb. 1999, JPL Pub. 99-17, pp. 217–227.
- [20] M. R. Azimi-Sadjadi and S. A. Zekavat, "Cloud classification using support vector machines," *Proc. IGARSS*, vol. 2, pp. 669–671, July 2000.
- [21] S. L. Tanimoto, *The Elements of Artificial Intelligence Using Common Lisp*. New York: W. H. Freeman, 1990.
- [22] S. Dellepiane and F. Fontana, "Extraction of intensity connectedness for image processing," *Pattern Recognit. Lett.*, vol. 16, no. 3, pp. 313–324, Mar. 1995.
- [23] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *J. ACM*, vol. 13, pp. 471–494, 1966.
- [24] R. Lumia, L. G. Shapiro, and O. Zuniga, "A new connected components algorithm for virtual memory computers," *Comput. Vis., Graph., Image Process.*, vol. 22, pp. 287–300, 1983.
- [25] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Upper Saddle River, NJ: Prentice-Hall, 2001, pp. 56–63.
- [26] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. B. Mitchell, "An efficiently computable metric for comparing polygonal shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 209–216, Mar. 1991.
- [27] I. Tomek, "Two algorithms for piecewise-linear continuous approximation of functions of one variable," *IEEE Trans. Comput.*, vol. C-23, no. 4, pp. 445–448, 1974.
- [28] K. Wall and P. E. Danielsson, "A fast sequential method for polygonal approximation of digitized curves," *Comput. Vis., Graph., Image Process.*, vol. 28, pp. 220–227, 1984.
- [29] D. Magee and R. Boyle, "Building shape models from image sequences using piecewise linear approximation," in *Proc. Brit. Machine Vision Conf.*, Southampton, U.K., Sept. 1998, pp. 398–408.
- [30] R. Chen, M. Ouyang, and P. Revesz, "Approximation data in constraint databases," in *Proc. 4th Int. Symp. Abstraction, Reformulation and Approximation*, TX, July 2000, pp. 124–143.
- [31] P. Revesz, *Introduction to Constraint Databases*. Berlin, Germany: Springer-Verlag, 2001, pp. 250–263.
- [32] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [33] C. Chang and C.-J. Lin. (2001) LIBSVM: A library for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [34] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," MIT, Cambridge, MA, Tech. Rep. A.I. Memo 1599, Dec. 1996.
- [35] T. Joachims, "Estimating the generalization performance of a SVM efficiently," in *Proc. 7th Int. Conf. Machine Learning*, Stanford, CA, June 29–July 2 2000, pp. 431–438.
- [36] V. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Comput.*, vol. 12, no. 9, pp. 2013–2036, 2000.
- [37] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, no. 1, pp. 131–159, 2002.
- [38] O. Chapelle and V. Vapnik, "Model selection for support vector machines," *Adv. Neural Inform. Process. Syst.*, vol. 12, 1999.
- [39] J.-H. Lee and C.-J. Lin. (2003) Automatic model selection for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/modelselect.ps.gz>.
- [40] S. D. Kim, J.-H. Lee, and J. K. Kim, "A new chain-coding algorithm for binary images using run-length codes," *Comput. Vis., Graph., Image Process.*, vol. 41, pp. 114–124, 1988.

- [41] H. Freeman, "Computer processing of line-drawings images," *ACM Comput. Surv.*, vol. 6, no. 1, pp. 57–95, 1974.
- [42] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 673–689, July 1996.
- [43] W. M. Alley, "The Palmer drought severity index: Limitations and assumptions," *J. Appl. Meteorol.*, vol. 23, pp. 1100–1109, 1984.

Jiang Li (S'02–M'03) received the B.S. and M.S. degrees in 1995 and 1998, respectively, both from the Beijing Institute of Technology, Beijing, China, and the Ph.D. degree from the University of Nebraska, Lincoln, in 2003.

He has been a Microsoft Certified System Engineer since 1998. From 1998 to 1999, he was an Instructor and Research and Development Engineer with Digital Audio and Video Engineering Center at the Beijing Broadcasting Institute, Beijing, China. He is working on the development of intelligent

image mining techniques to identify hidden patterns in remotely sensed images. He has published several papers in refereed journals and has presented several conference papers. He is currently an Assistant Professor with the Department of Computer Science and Information Technology at Austin Peay State University, Clarksville, TN.

Ram M. Narayanan (S'84–M'88–SM'91–F'01) received the B.Tech. degree from the Indian Institute of Technology, Madras, India, in 1976, and the Ph.D. degree from the University of Massachusetts, Amherst, in 1988, both in electrical engineering.

During 1976–1983, he was employed as a Research and Development Engineer at Bharat Electronics Ltd., Ghaziabad, India. He joined the University of Nebraska, Lincoln, in 1988 where he is currently a Professor of electrical engineering.

At the University of Nebraska, he established the Environmental Remote Sensing Laboratory that conducts research on remote sensing science and technology development. His research interests lie in the area of radar and laser remote sensing of natural surfaces and the development of remote sensing systems for environmental, agricultural, and meteorological applications. He is especially interested in the development of airborne remote sensing techniques. He has authored or coauthored over 50 journal papers, as well as over 100 conference publications.

Dr. Narayanan is active in the IEEE Geoscience and Remote Sensing Society. He served as Newsletter Editor during 1992–1994. He led the development of the successful proposal to host the International Geoscience and Remote Sensing Symposium (IGARSS '96) in Lincoln, NE, and served as the General Chairman for the symposium. He was elected to serve in the IEEE GRSS Administrative Committee for the period 1995–2000. He is a member of the American Society for Photogrammetry and Remote Sensing (ASPRS), International Union of Radio Science (URSI) Commission F, International Electromagnetics Academy (IEA), and American Institute of Aeronautics and Astronautics (AIAA). He has won awards for teaching, research, and service at the University of Nebraska and was named the Outstanding Faculty Member by the IEEE Nebraska Student Section in 1990–1991. He was an invited participant in a National Science Foundation workshop on the application of remote sensing technology for monitoring power and dispersed civil infrastructure systems in April 1994. He served on the National Research Council's Committee on Radio Frequencies during the period 1997–2000. He currently serves as the Scientific Advisor to the joint NASA/DoT Program in Remote Sensing.