DATA COMPRESSION AND CRITICAL POINTS DETECTION

USING NORMALIZED SYMMETRIC SCATTERED MATRIX

Khagendra Thapa B.Sc. B.Sc(Hons) CNAA, M.Sc.E. M.S. Ph.D. Department of Surveying and MappingFerris State University Big Rapids, Michigan 49307

BIOGRAPHICAL SKETCH

Khagendra Thapa is an Associate Professor of Surveying and Mapping at Ferris State University. He received his B.SC. in Mathematics, Physics, and Statistics from Tribhuvan University Kathmandu, Nepal and B.SC.(Hons.) CNAA in Land Surveying from North East London Polytechnic, England, M.SC.E. in Surveying Engineering from University of New Brunswick, Canada and M.S. and Ph.D. in Geodetic Science from The Ohio State University. He was a lecturer at the Institute of Engineering, Kathmandu Nepal for two years. He also held various teaching and research associate positions both at The Ohio State University and University of New Brunswick.

ABSTRACT

The problems of critical points detection and data compression are very important in computer assisted cartography. In addition, the critical points detection is very useful not only in the field of cartography but in computer vision, image processing, pattern recognition, and artificial intelligence. Consequently, there are many algorithms available to solve this problem but none of them are considered to be satisfactory. In this paper, a new method of finding critical points in digitized curve is explained. This technique, based on the normalized symmetric scattered matrix is good for both critical points detected by this algorithm are compared with those detected by humans.

INTRODUCTION

The advent of computers have had a great impact on mapping sciences in general and cartography in particular. Now-a-days-more and more existing maps are being digitized and attempts have been made to make maps automatically using computers. Moreover, once we have the map data in digital form we can make maps for different purposes very quickly and easily. Usually, the digitizers tend to digitize more data than what is required to adequately represent the feature. Therefore, there is a need for data compression without destroying the character of the feature. This can be achieved by the process of critical points detection in the digital data. There are many algorithms available in the literature for the purpose of critical points detection. In this paper, a new method of critical points detection is described which is efficient and has a sound theoretical basis as it uses the eigenvalues of the Normalized Symmetric Scattered (NSS) matrix derived from the digitized data.

DEFINITION OF CRITICAL POINTS

Before defining the critical points, it should be noted that critical points in a digitized curve are of interest not only in the field of cartography but also in other disciplines such as Pattern Recognition, Image Processing, Computer Vision, and Computer Graphics. Marino (1979) defined critical points as "Those points which remain more or less fixed in position, resembling a precis of the written essay, capture the nature or character of the line".

In Cartography one wants to select the critical points along a digitized line so that one can retain the basic character of the line. Researchers both in the field of Computer Vision and Psychology have claimed that the maxima, minima and zeroes of curvature are sufficient

to preserve the character of a line. In the field of Psychology, Attneave (1954) demonstrated with a sketch of a cat that the maxima of curvature points are all one needs to recognize a known object. Hoffman and Richards (1982) suggested that curves should be segmented at points of minimum curvature. In other words, points of minimum curvature are the critical points. They also provided experimental evidence that humans segmented curves at points of curvature minima. Because the minima and maxima of a curve depend on the orientation of the curve, the following points are considered as critical points:

- 1. curvature maxima
- 2. curvature minima
- 3. end points
- 4. points of intersection.

It should be noted that Freeman (1978) also includes the above points in his definition of critical points.

Hoffman and Richards (1982) state that critical points found by first finding the maxima, minima, and zeroes of curvature are invariant under rotations, translations, and uniform scaling. Marimont (1984) has experimentally proved that critical points remain stable under orthographic projection.

The use of critical points in the fields of Pattern Recognition, and Image Processing has been suggested by Brady (1982), and Duda and Hart (1973). The same was proposed for Cartography by Solovitskiy (1974), Marino (1978), McMaster (1983), and White (1985).

Importance of Critical Point Detection in Line Generalization

Cartographic line generalization has hitherto been a subjective process. When one wants to automate a process which has been vague and subjective, many difficulties are bound to surface. Such is the situation with Cartographic line generalization. One way to tackle this problem would be to determine if one can quantify it (i.e. make it objective) so that it can be solved using a digital computer. Many researchers such as Solovitskiy (1974), Marino (1979), and White (1985) agree that one way to make the process of line generalization more objective is to find out what Cartographers do when they perform line generalization by hand? In addition, find out what in particular makes the map lines more informative to the map readers. Find out if there is anything in common between the map readers and map makers regarding the interpretation of line character.

Marino (1979) carried out an empirical experiment to find if Cartographers and noncartographers pick up the same critical points from a line. In the experiment, she took different naturally occurring lines representing various features. These lines were given to a group of Cartographers and a group of non-cartographers who were asked to select a set of points which they consider to be important to retain the character of the line. The number of points to be selected was fixed so that the effect of three successive levels or degrees of generalization could be detected. She performed statistical analysis on the data and found that cartographers and non-cartographers were in close agreement as to which points along a line must be retained so as to preserve the character of these lines at different levels of generalization. When one says one wants to retain the character of a line what he/she really means is that he/she wants to preserve the basic shape of the line as the scale of representation decreases. The purpose behind the retention of the basic shape of the line is that the line is still recognized as a particular feature- river, coastline or boundary despite of the change in scale. The assumption behind this is that the character of different types of line is different. That is to say that the character of a coastline is different from that of a road. Similarly, the character of a river would be different from that of a transmission line and so on.

The fact that during the process of manual generalization one retains the basic shape of the feature has been stated by various veteran Cartographers. For example, Keates (1973) states, "... each individual feature has to be simplified in form by omitting minor irregularities and retaining only the major elements of the shape". Solovitskiy (1974) identified the following quantitative and qualitative criteria for a correct generalization of lines:

- 1. The quantitative characteristics of a selection of fine details of a line.
- 2. Preservation of typical tip angles and corners
- 3. Preservation of the precise location of the basic landform lines.
- 4. Preservation of certain characteristic points.5. Preservation of the alternation frequency and specific details.

He further states "The most important qualitative criteria are the preservation of the general character of the curvature of a line, characteristic angles, and corners...". In the above list, what Solovitskiy is basically trying to convey is that he wants to retain the character of a feature by preserving the critical points. Buttenfield (1985) also points out the fact that Cartographers try to retain the basic character of a line during generalization. She states "... Cartographer's attempt to cope objectively with a basically inductive task, namely, retaining the character of a geographic feature as it is represented at various Cartographic reductions".

Boyle (1970) suggested that one should retain the points which are more important (i.e. critical points) during the process of line generalization. He further suggested that these points should be hierarchical and should be assigned weights (1-5) to help Cartographers decide which points to retain.

Campbell (1984) also observes the importance of retaining critical features. He states, "One means of generalization involves simply selecting and retaining the most critical features in a map and eliminating the less critical ones". The fact that retention of shape is important in line generalization is also included in the definition of line generalization. The DMA (Defense Mapping Agency) definition states as "Smoothing the character of features without destroying their visible shape". Tobler as referenced in Steward (1974) also claims that the prime function of generalization is "... to capture the essential characteristics of ... a class of objects, and preserve these characteristics during the change of scale".

Advantages of Critical Point Detection

According to Pavlidis and Horowitz (1974), Roberge (1984), and McMaster (1983) the

detection and retention of critical points in a digital curve has the following advantages:

- 1. Data compaction as a result plotting or display time will be reduced and less storage will be required.
- 2. Feature extraction.
- 3. Noise filtering.
- 4. Problems in plotter resolution due to scale change will be avoided.
- 5. Quicker vector to faster conversion and vice-versa.
- 6. Faster choropleth shading. This means shading color painting the polygons.

Because of the above advantages, research in this area is going on in various disciplines such as Computer Science, Electrical Engineering, Image Processing, and Cartography.

Literature Review

The proliferation of computers not only has had a great impact on existing fields of studies but also created new disciplines such as Computer Graphics, Computer Vision, Pattern Recognition, Image Processing, Robotics etc. Computers play an ever increasing role in modern day automation in many areas. Like many other disciplines, Mapping Sciences in general and Cartography in particular have been greatly changed due to the use of computers. It is known from experience that more than 80% of a map consists of lines. Therefore, when one talks about processing maps, one is essentially referring to processing lines. Fortunately, many other disciplines such as Image Processing, Computer Graphics, and Pattern Recognition are also concerned with line processing. They might be interested in recognizing shapes of various objects, industrial parts recognition, feature extraction or electrocardiogram analysis etc.

Whatever may be the objective of line processing and whichever field it may be, there is one thing in common viz: it is necessary to retain the basic character of the line under consideration. As mentioned above one needs to detect and retain the critical points in order to retain the character of a line. There is a lot of research being carried out in all the above disciplines as to the detection of critical points. Because the problem of critical points detection is common to so many disciplines, it has many nomenclatures. A number of these nomenclatures (Wall and Danielson, 1984), (Dunham, 1986), (Imai and Iri, 1986), (Anderson and Bezdek, 1984), (Herkommer, 1985), (Freeman and Davis, 1977), (Rosenfeld and Johnston, 1973), (Rosenfeld and Thurston, 1971), (Duda and Hart, 1973), (Opheim, 1982), (Williams, 1980), (Roberge, 1984), (Pavlidis and Horowitz, 1974), (Fischler and Bolles, 1983,1986), (Dettori and Falcidieno, 1982), (Williams, 1978) are lised below:

- 1. Planer curve segmentation
- 2. Polygonal Approximation
- 3. Vertex Detection
- 4. Piecewise linear approximation
- 5. Corner finding
- 6. Angle detection
- 7. Line description

- 8. Curve partitioning
- 9. Data compaction
- 10. Straight line approximation
- 11. Selection of main points
- 12. Detection of dominant points
- 13. Determination of main points.

Both the amount of literature available for the solution of this problem and its varying nomenclature indicate the intensity of the research being carried out to solve this problem. It is recognized by various researchers (e.g. Fischler and Bolles, 1986) that the problem of critical points detection is in fact a very difficult one and it still is an open problem. Similarly, the problem of line generalization is not very difficult if carried out manually but becomes difficult if one wants to do it by computer. Because of the subjective nature of this problem and due to the lack of any criteria for evaluation of line generalization, it has been very difficult to automate this process. Recently some researchers for example (Marino, 1979) and (White, 1985) have suggested that one should first find critical points and retain them in the process of line generalization.

Algorithms for Finding Critical Points

As noted in the previous section, there are many papers published on critical points detection which is identified by different names by different people. It should, however, be noted that the detection is not generic but, as indicated by Fischler and Bolles (1986) depends on the following factors:

- 1. purpose
- 2. vocabulary
- 3. data representation
- 4. past experience of the 'partitioning instrument'. In cartography it would mean the past experience of the cartographer.

It is interesting to note that the above four factors are similar to the controls of line generalization that Robinson et al. (1985) have pointed out. However, the fourth factor viz: past experience and mental stability of the Cartographer is missing from the latter list.

THE NATURE OF SCATTER MATRICES AND THEIR EIGENVALUES

Consider the geometry of the quadratic form associated with a sample covariance matrix. suppose $P = (p_1, p_2, \dots, p_n)$ be a finite data set in R^2 and P is a sample of n independently and identically distributed observations drawn from real two dimensional population.

Let $(\mu, _)$ denote the population mean vector and variance matrix and let (v_p, V_p) be the corresponding sample mean vector and sample covariance matrix these are then given by (Uotila, 1986)

$$v_{p} = \sum p_{i}/n; \quad V_{p} = \sum (p_{i} - v_{p}) (p_{i} - v_{p})$$
(1)

Multiply both sides of the equation for V_n by (n - 1) and denote the RHS by S_n viz:

The matrices S_p and V_p are both 2x2 symmetric and positive semi-definite. Since these matrices are multiples of each other they share identical eigen-spaces.

According to Anderson and Bezdek (1983) one can use the eigenvalue and eigenvector structure of S_p to extract the shape information of the data set it represents. This is because the shape of the data set is supposed to mimic the level shape of the probability density function f(x) of x. For example, if the data set is bivariate normal, S_p has two real, non-negative eigenvalues. Let these eigenvalues be λ_1 and λ_2 . Then the following possibilities exist (Anderson and Bezdek, 1983):

1. If both λ_1 and $\lambda_2 = 0$, then the data set P is degenerate, and S_p is invertible and there exist with probability 1, constants a, b, and c such the ax + by + c = 0. In this case the sample data in P lie on a straight line.

2. If $\lambda_1 > \lambda_2 > 0$, then the data set represent an elliptical shape.

3. If $\lambda_1 = \lambda_2 > 0$, then the sample data set in P represent a circle.

EIGENVALUES OF THE NORMALIZED SYMMETRIC SCATTER MATRIX (NSS)

Supposing that one has the following data:

$$P = (P_1, P_2, \dots, p_n)$$

where $P_1 = (x_1, y_1)$

Then the normalized scattered matrix A is defined as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = S_p / \text{trace} (S_p)$$
(3)

For the above data set A is given by:

$$Deno = \sum ((x_{1} - x_{m})^{**2} + (y_{i} - y_{m})^{**2})$$
(4)

$$a_{11} = 1/Deno \sum (x_{1} - x_{m})^{**2}$$
(3)

$$a_{12} = 1/Deno \sum (x_{1} - x_{m})(y_{i} - y_{m})$$
(4)

$$a_{21} = 1/Deno \sum (x_{1} - x_{m})(y_{i} - y_{m})$$
(5)

$$a_{22} = 1/Deno \sum (y_{1} - y_{m})^{**2}$$
(5)

where $v_{x} = (x_{m}, y_{m})$ is the mean vector defined as

$$x_m = \sum x_n/n$$
, and $y_m = \sum y_n/n$ (6)

Note that the denominator in (3) will vanish only when all the points under consideration are identical.

The characteristic equation of A is given by:

$$|\mathbf{A} - \boldsymbol{\lambda} \mathbf{I}| = 0 \tag{7}$$

which may be written as (for 2x2 matrix)

$$|\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}| = 0 \tag{7}$$

which may be written as (for 2x2 matrix)

$$\lambda^2$$
 - trace(A) λ + Det (A) = 0

where Det(A) = Determinant of A.

By design the trace of A is equal to 1. Hence the characteristics equation of A reduces to

(8)

$$\lambda^2 + \operatorname{Det}(A) = 0 \tag{9}$$

The roots of this equation are the eigenvalues and are given by:

$$\lambda_1 = (1 + \sqrt{1 - 4^* \text{Det}(A)})/2 \text{ and } \lambda_2 = (1 - \sqrt{1 - 4^* \text{Det}(A)})/2$$
 (10)

For convenience put $D_x = \sqrt{1 - 4^* \text{Det}(A)}$, then

$$\boldsymbol{\lambda}_1 = (1 + \mathbf{D}_x)/2 \tag{11}$$

$$\lambda_2 = (1 - D_x)/2 \tag{12}$$

Now 1 and 2 satisfy the following two conditions:

$$\lambda_1 + \lambda_2 = 1 \tag{13}$$

Since the sum of the roots of an equation of the form

$$ax^2 + bx + c = 0$$
 are $\lambda_1 + \lambda_2 = -b/a$

Subtracting (12) from (11), one obtains

$$\lambda_1 - \lambda_2 = D_x \tag{14}$$

Since the eigenvalues λ_1 and λ_2 satisfy the equations (13) and (14) the three cases discussed previously reduce to the following from (Anderson and Bezdek, 1983):

1. The data set represent a straight line if and only if $D_x = 1$

2. The data set represent an elliptical shape if and only if $0 \le D \le 1$

3. The data set represent a circular shape if $D_{x} = 0$.

ALGORITHM TO DETECT CRITICAL POINTS USING NSS MATRIX

The fact that the analysis of the eigenvalues of the NSS matrix can be used to extract shape of the curve represented by the data set, may be exploited to detect critical points in the digital curve. Assuming that the data is gross error free, and devoid of excessive noise, one can outline the algorithm to detect critical points in the following steps:

- 3. If D_x is greater than a certain tolerance (e.g. 0.95) add one more point to the data and repeat from step 2.
- 4. If D_x is less than the tolerance point, point 2 is a critical point. Retain point 2 and repeat the process from step 1 with point two as the new starting point.
- 5. Repeat the process until the end of the data set is reached.

Results of Critical Points Detection by NSS Matrix

The algorithm discussed in the previous section is useful in detecting the critical points in vector data. The only parameter involved in this technique is D_x which was defined earlier. by varying the value of D_x between say 0.8 to 1.0 one can get a varying amount of detail in a curve. Figure 1 shows the selected critical points for the test figure for $D_x = 0.96$.

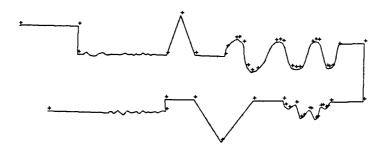


Figure 1: Results of Critical Points Selection by NSS Matrix. There were 50 points selected.

There are 50 points selected in this figure. It is clear from the figure that this method will be very useful for compression of digitized data since it retains the overall shape of the curve without retaining the unnecessary points.

COMPARISON BETWEEN MANUAL AND ALGORITHMIC CRITICAL POINTS DETECTION

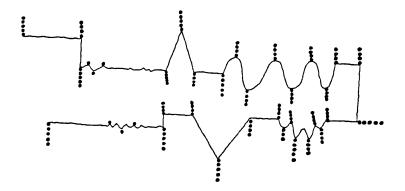
In this section, results of critical points detection in the test figure by a group of people are given. These results are then compared with the results obtained from the NSS matrix technique of critical points detection.

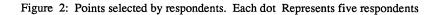
MANUAL CRITICAL POINTS DETECTION: THE EXPERIMENT

In order to find if the NSS matrix method critical points detection can mimic humans or not, the test figure was given to a group of 25 people who had at least one course in Cartography.

In addition, they were told about the nature of critical points. In the experiment, they were asked to select not more than 50 points from the test figure. The results of critical points detection by the above group are shown in figure 2.In figure 2 each dot represents 5 respondents. A point was rejected if it was selected by less than four respondents.

A comparison of figures 1 and 2 reveals that the critical points selected by the humans are almost identical with those selected by NSS matrix algorithm. Only five or less people slightly disagreed in the selection of a few points with NSS matrix algorithm. It should be pointed out that the results of critical points selection could have been different if the respondents were asked to select all the critical points in the curve. However, the NSS matrix algorithm also can be made to detect different levels of critical points by simply changing the value of the parameter D₂.





CONCLUSION

1. The analysis of the eigenvalues of the normalized symmetric scattered matrix provides an useful way of detecting critical points in digitized curves. Consequently, this technique may be used for data compression for digitized curves.

2. The NSS matrix algorithm for critical points detection can mimic humans in terms of critical points detection in a digitized curve.

ACKNOWLEDGEMENT

I express my gratitude to Dr. Toni Schenk for providing partial funding for this research under the seed grant 22214 and NASA project 719035.

REFERENCES

- Anderson, I.M. and J.C. Bezdek (1984), "Curvature and Tangential Deflection of Discrete Arcs: A Theory Based on the Commutator of Scatter Matrix Pairs and Its Application to Vertex Detection in Planer Shape Data", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, NO. 1, pp. 27-40.
- Attneave, F. (1954), "Some Informational Aspects of Visual Perception", Psychological Review, Vol. 61, pp. 183-193.
- Boyle, A.R. (1970), "The Quantized Line", The Cartographic Journal, Vol. 7, No. 2, pp. 91-94.
- Buttenfield, B. (1985), "Treatment of the Cartographic Line", Cartographica, Vol. 22, No. 2, pp. 1-26.
- Campbell, J. (1984), Introductory Cartography, Prentice Hall, Inc., Englewood Cliffs, NJ 07632.
- Davis, L.S. (1977), "Understanding Shape: Angles and Sides", IEEE Transactions on Computers, Vol. C-26, No.3, pp. 236-242.
- Dettori, G. and B. Falcidieno (1982), "An Algorithm for Selecting Main Points on a Line," Computers and Geosciences, Vol. 8, pp.3-10.
- Douglas, D.H. and T.K. Peucker (1973), "Algorithms for the Reduction of the Number of points Required to Represent a Digitized Line for its Character", The Canadian Cartographer Vol. 10.
- Duda, R.O. and P.E. Hart (1973), Pattern Classification and Scene Analysis, Willey Interscience.
- Dunham, J.G. (1986), "Optimum Uniform Piecewise Linear Approximation of Planer Curves", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 1.
- Fischler, M.A. and R.C. Bolles (1986), "Perceptual Organization and Curve Partitioning," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol PAMI-8, No. 1.
- Freeman, H. (1978), "Shape Description Via the Use of Critical Points," Pattern Recognition, Vol. 10, pp. 159-166.
- Herkommer, M.A. (1985), "Data-Volume Reduction of Data Gathered along Lines Using the Correlation Coefficient to Determine Breakpoints," Computers and Geos ciences, Vol. 11, No. 2, pp. 103-110.

- Hoffman, D.D. and W.A. Richards (1982), "Representing Smooth Plane Curves for Recognition: Implications for Figure-Ground Reversal," Proceedings of the National Conference on Artificial Intelligence," Pittsburgh PA, pp. 5-8.
- Imai, H. and M. Iri (1986), "Computational Geometric Methods for Polygonal Approxima tions of a Curve," Computer Vision Graphics and Image Processing, Vol. 36, pp. 31-41.
- Keates, J.S. (1981), Cartographic Design and Production, Thetford, Great Britain: Longman.
- Marino, J. S. (1979), "Identification of Characteristics Points along Naturally Occuring Lines: An Empirical Study," The Canadian Cartographer, Vol. 16, No. 1, pp. 70-80.
- Marino, J. (1978), "Characteristics Points and Their Significance to Line Generalization," Unpublished M.A. Thesis University of Kansas.
- McMaster, R.B. (1983), "A Quantative Analysis of Mathematical Measures in Linear Simplification," Ph.D. Dissertation Dept. of Geography-Meteorology, University of Kansas, Kansas City.
- Opheim, H. (1982), "Fast Data Reduction of a Digitized Curve," Geo-Processing, Vol. 2, pp. 33-40.
- Ramer, Urs (1972). "An Iterative Procedure for the Polygonal Approximation of Plane Curves," Computer Graphics and Image Processing, Vol. 1, No. 3, pp. 244-256.
- Reumann, K. and A. Witkam (1974), "Optimizing Curve Segmentation in Computer Graphics," International Computing Symposium, Amsterdam, Holland, pp. 467-472.
- Roberge, J. (1985), "A Data Reduction Algorithm for Planer Curves, " Computer Vision Graphics and Image Processing, Vol. 29, pp. 168-195.
- Robinson, A. H. and B.B. Petchenik (1976), "The Nature of Maps," University of Chicago Press, Chicago.
- Rosenfeld, A. and E. Johnston (1973), "Angle Detection on Digital Curves, " IEEE Transactions on Computers, Vol. C-22, pp. 875-878.
- Sklansky, J. and V. Gongalez (1980), "Fast Polygonal Approximation of Digitized Curves," Pattern Recognition, Vol. 12, pp. 327-331.
- Solovitskiy, B.V. (1974), "Some Possibilities for Automatic Generalization of Outlines," Geodesy, Mapping and Photogrammetry, Vol. 16, No.3.
- Spath, H. (1974), Spline Algorithms for Curves and Surfaces Unitas Mathematica

- Publication, Winnapeg, Translated from German by W.D. Hoskins and H.W. Sagar. Steward, H.J. (1974), Cartographic Generalization: Some Concepts and Explanation, Cartographica Monograph No. 10., University of Toronto Press.
- Sukhov, V. I. (1970), "Application of Information Theory in Generalization of Map Contents," International Yearbook of Cartography, Vol. 10, pp. 48-62.
- Thapa, K. (1987), "Critical Points Detection: The First Step To Automatic Line Generali zation." Report Number 379 The Department of Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio.
- Uotila, U.A. (1986), "Adjustment Computation Notes," Dept. of Geodetic Science and Surveying, The Ohio State University.
- Wall, K. and P. Danielsson (1984), "A Fast Sequential Method for Polygonal Approximat tion of Digitized Curves," Computer Vision Graphics and Image Processing, Vol. 28, pp. 220-227.
- White, E.R. (1985), "Assessment of Line Generalization Algorithms Using Characteristic Points," The American Cartographer Vol. 12, No. 1.
- Williams, C.M. (1981), "Bounded Straight Line Approximation of Digitized planer Curver and Lines, "Computer Graphics and Image Processing, Vol. 16, pp. 370-381.