A Finding Method of Shape Theme Based on Wavelet Time Series

Guo Qiu
College of Software,
University of Science and Technology Liaoning,
AnShan，Liaoning，114051， China
E-mail:asguoqiu@126.com

Guo Lu
Colege of Science and Technology,
Shenyang Agricultural University ,
ShenYang, Liaoning,110161, China
E-mail:asguolu@126.com

***Abstract*—Finding shape theme has raised great attention in the database of shapes. According to the problem of incompatible about accuracy and complexity in the shape theme search algorithm ,this paper proposed a finding theme algorithm using the multi-resolution analysis of wavelet and the processing capability of reduction dimension of time sequence , accurately calculated the similarity between different objects combining with the Euclidean distance formula, and achieved satisfactory results. Through the comparison between the real data sets to test and traditional shape theme algorithm, it shows that the method has good stability and reliability, and ensure the real-time processing ability of the closed contour shape's overall matching.**

Keywords-time series; shape matching; Haar wavelet; Euclidean distance

I. INTRODUCTION

When observating environment, people first notice the object and the surrounding environment of the color, texture, shape and space relations and so on. The shape of objects is one of the most basic characteristics of the feeling. The figure is a very important kind of data object, basing on the graphic data analysis can reveal the change of things and the law of development, and provide a basis for the later scientific decision-making. The pattern which repeated in the database is called theme, in recent years it has become an important issue to find topic in the shape of a database, the shape can be applied in the field of anthropology, manuscript identification, cultural identification and so on[1].

In the past ten years, how to effective orientation the image that similar to target image has attracted much attention in the image processing field [2~4]. In recent years, many scholars have done a lot of work in database query shape theme. According to projection mapping method [5], Jeremy Buhler and Martin Tompa proposed using random mapping method to inquiry; and then Eamonn Keogh, Li Wei proposed using SAX random projection method for shape theme queries. The second topic query method is to scanning boundary of its graphical shapes by using graphics centroid , then get the distance between them and convert to time series data, then using the method of SAX to compress it [6~7], in order to reduce the dimension of data, and then according to the similarity comparison of time sequence, put the similar graphics into a same kind theme. The above method mostly adopts a numerical or a feature vector to represent the target, it is difficult to express fully feature of target shape and lost many data points of the original shape, so the accuracy of query results will be affected.

This paper proposed a method based on wavelet analysis and time series in the shape of subject searching method. First changed the goal of contour shape as time series form, then, using Haar wavelet multi resolution characteristics to fast reduce the dimension of time series, it reduced the matching space very well. Haar wavelet also has very strong pruning capabilities, it introduced noise very well through reduced dimensionality, cut off those who may not match the sequence, and not be missed when generate the matching, the accuracy of query results can be greatly improved.

II. BASED ON WAVELET TIME SEQUENCE SHAPE THEME SEARCHING METHOD

A. Related Concepts

Time series: time series is the real set which arranged according to the chronology order.

Pattern matching: for the time series of two graphs T1 , T2 and a limit value M>0, if the distance between T1 and T2 is smaller than M, then T1 matched T2.

Threshold: also called critical value, it refers a minimum or maximum value which an effect produced. This paper refers the highest value that the two shapes matched.

The identified of Target shape centroid: determined the centroid of the shape by applying the Green's theorem [8]. Calculated the area before determined the shape centroid, according to the equation (1).

$$\bar{x} = \frac{\mu_x}{A} \qquad \mu_x = \frac{1}{6}\sum_{i=0}^{n-1}(x_{i+1} + x_i)a_i \qquad A = \frac{1}{2}\sum_{i=0}^{n-1} a_i$$

$$\bar{y} = \frac{\mu_x}{A} \qquad \mu_y = \frac{1}{6}\sum_{i=0}^{n-1}(y_{i+1} + y_i)a_i \qquad a_i = x_i y_{i+1} - x_{i+1} y_i \quad (1)$$

The identified of Time sequence: distributed the outline of shape by the fixed angle according to the centroid , then sequence scanning through clockwise or counterclockwise , formed a time series from the distance between the mass center of shape and the outline point which gets from the average distribution [9]. As shown in figure 1.

Published by Atlantis Press, Paris, France.

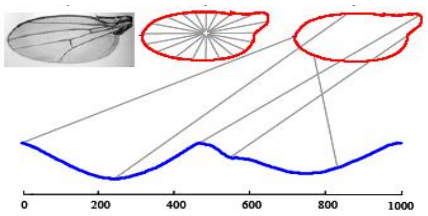

Figure 1. The convert from shape to time sequence

B. Similarity Measure

After obtaining time sequence of each target shape, it need to calculate the similarity of these shape time series to get the conclusion.

The method about calculating time sequence similarity described as follows:

- calculated the distance between a target shape time series and another.

$$D(Q,C) = \sqrt{\sum_{i=1}^{n} (q_i - c_i)^2} \qquad (2)$$

  Then compared them with a threshold value, the matching results is the one that less than the threshold. The smaller the distance value, the higher the matching precision.
- Arrayed the distance from small to large ,then got the similarity from large to small, and obtained the best matching result combined with the shape itself.

The method express the similarity of two target by a value which is not a vector, based on the calculation results, the greater results is the less similar , the smaller results the more similar.

C. Haar Wavelet Method

The method of wavelet analysis is a kind of general digital signal processing technology. Similar to the Fu Liye transform, wavelet analysis according to the input analog, transform into a series of wavelet parameters, to restore the original signal approximation. There are many types Wavelet, the most common and most simple is Haar wavelet.

Discrete wavelet transform (DWT) express signal as approximate quantity (As) and the amount of details (Ds) which can localizate in the time domain and frequency domain, DWT is a discrete convolution process, as shown in formula (3):

$$w^* x_t = \sum_{i=-\infty}^{\infty} w_i x_{t-1} \qquad (3)$$

The x_t is the original signal, w is a low-pass or high-pass filtering related to the prototype wavelet, w_i is the nonlinear time series behavior which packaged by parameter set, in the wavelet decomposition ,different levels (i=1 to L) details Di can capture the local fluctuations in the entire time series.

Error tree can well describe the characteristics of Haar wavelet decomposition, hierarchical structure based on wavelet transform (can be created in linear time O *(N)*), each inner node Ci (i=0, ... 15) corresponds to a wavelet coefficients, each leaf node (i=0, ….., 15) corresponds to a value of the original data. One dimensional data vector A contains N=16 data values, such as [27,71,17,31,59,3, 43, 59100, 42, 0, 58, 30,88,72130], then gets the low scale data average in pair to express the average value [49,24,31,51, 71,29,59101], in which the average of 27 and 71 is 49, and so on. As shown in table 1.

TABLE I. THE CALCULATION OF ONE-DIMENSIONAL HAAR WAVELET TRANSFORM

Scale	Average value	Detail coefficient
4	*[27,71,17,31,59,3,43,59,100,42,0,58,30,88,72,130]*	
3	*[49,24,31,51,71,29,59,101]*	*[-22,-7,28,-8,29,-29,-29,-29]*
2	*[36.5,41,50,80]*	*[12.5,-10,21,-21]*
1	*[38.75,65]*	*[-2.25,-15]*
0	*[51.875]*	*[-13.125]*

Obviously, lots of information has been lost in the average treatment process , in order to retrieve the original data value and capture the missing information, it needs to preserve the detail coefficient. For the Haar wavelet ,these detail coefficients are very simple, ci= (d_i- d_{i+1}) /2= (d_i+d_{i+1}) /2 –d_{i+1}, the information is not lost during this process. So, it is easy to reconstruce the 16 value of original data, it can get complete wavelet decomposition by recursive applicating treatment about the pairwise average and calculating the difference.

So, reduced the original data double, and reduced the dimension of original time series. Obviously, the algorithm can improve the compression match time after remove the redundant data limit.

D.Algorithm Description

The algorithm of a finding method of shape theme based on wavelet time series described as shown in figure 2.

```
Algorithm：MATCH_SELECT
Input: time sequence set C, threshold max_error
Output: matching the time series of result set Result_Matching_Pairs
1.While(C)  SeQuence_Count++;  // SeQuence_Count recording time sequence number
2.Pick_A_Sequence(T);  // pulls out a time series T f rom the time series of the set C
3.for(i = 0;i< length(T);i++)
4. Sequence_Array_2D[SeQuence_Count][i] = T[i];// put each time sequence in the array and store Sequence_Array_2D
5.Result_Matching_Pairs = ø;// initialization time sequence matching ,the result set is empty
6.Haar_Transform(Sequence_Array_2D, Haar_Array_3D);
                // using wavelet transform method to convert
7.for(i=0;i< SeQuence_Count -1;i++)
8. for(j = i+1;j < SeQuence_Count;j++)
9.  k=max_line;// max_line is the maximum number of lines of time sequence after the wavelet transform
10. While(k>-1)
11.  if(Distance(Haar_Array_3D[i][k], Haar_Array_3D[j][k])< max_error)
12.  {// time sequence of i and j is matching if the line k of the Euclidean distance of the time sequence i and j is less than a threshold,

13.  Result_Matching_Pairs←(i,j);// deposit the time sequence query results into
                The Result_Matching_Pairs result set
14.  Output_Matching_Pair(i,j); //output timing on i,j
15.  Break; }
```

Figure 2. The matching algorithm based on wavelet time series

In this algorithm, the line between 1 and 4 is a process which extracts time series from the time sequence set and saves them in a 2 dimensional array. The fifth line is the wavelet transform process, and converts the 2 dimensional array into 3 dimensional array. The line between 5 and 14 is the process of finding subject. Distance is the Euclidean distance function, the line between 9 and 13 judges whether the Euclidean distance of the time sequence i and j in line k is below threshold, if less than ,it means the time series i and j matching, otherwise need to continue calculate time series i and j in line k-1, and judge whether the Euclidean distance below the threshold, ... Until after the calculation of original time series (the zeroth line) Euclidean distance.

III.EXPERIMENTAL RESULTS AND ANALYSIS

The data set chicken used in the experiment is the data set which used by the expert Dr.Keogh in UCR University at November 15, 2006, and it can be download at www.cs.ucr.edu/~xxi/SDM07. The experiment used 1200 groups form data, each group with 2200 data points, we experiment with 100 groups of data, use 4 group as an example. As shown in Figure 3 ,four chickens shape (complete) contour curve C, C1, C2, C3. C is a reference shape, C1, C2, C3 are waiting for recongnition shape. We can see from the graph, C and C1, C2 shape are most similar ,they can be classified as a category topic matching, while C3 and the others are not close, so cannot be classified as a category. The shape functions are shown in the figure a to d, which the smaller the closed contour curves of the various types of wavelet feature matching degree is, the better the reliability of this query method is.

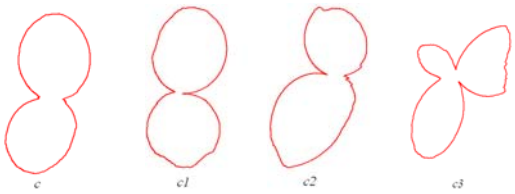

Figure 3. The simulation of four chickens shape contour curve

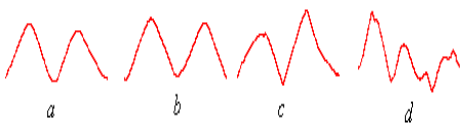

Figure 4. The corresponding time sequence of the simulation

The algorithm using VC++ achieve, in addition, from the results of the experiment indicated the operation speed of model, the existing Pentium PC machine basically meet the needs of real time processing. Using the above data set, we do experiment on four kinds of shape theme inquiry method , the brute force method, random projection method, sax projection method and wavelet transform method. Comparison of the accuracy of four kind of method in the same threshold, the experimental results as shown in figure 5:

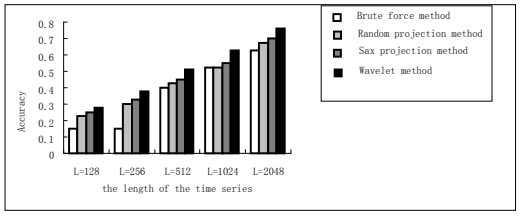

Figure 5. the query results of the shape data sets for different sequence length

Fig 5 expressed in the same threshold conditions, the query results of the query sequence of the shape data with different length. In Fig 5 irrespective the number of the sequence length, the query accuracy of wavelet method is better than the other three algorithms. From the experiment comparison, obtained different values of accuracy in different length of the sequence, the value of precision is increased with the the sequence length increasing.

The above experimental results show that using Haar wavelet method is better than the other three matching method, the query result is consistently the highest. One of the reasons is that the multi resolution characteristics of Haar wavelet can fast reduce the dimensionality of time series, can reduce the matching space very well.The second reason is that the Haar wavelet has a very strong pruning capability, it can dispose of the noise very well through reducing dimensionality, and cut off those sequence which may not match, and can't generate the missing when matched.

SUMMARY

Proposed the shape topic search algorithm based on wavelet analysis and time series. Using the multi-resolution characteristics of wavelet, the algorithm can compress the time series very well, better reflect the similarity in combination with the Euclidean distance formula by calculating the similarity between the shapes, and improve the query precision, achieve a better matching effect by using the method in the whole sequence matching. Experimental results show that the method further improve the precision of shape theme queries, reduce the computational cost and better than the existing methods. This method is simple and easy to realize test, and can be applied to many fields. The next phase of work is to explore the use of other orthogonal wavelet, prove its feasibility and compare the similarity matching method with Haar wavelet.

REFERENCES

[1] Xiaopeng Xi, Eamonn Keogh, Li Wei, et al. Finding Motifs in Database of Shapes[C].In Proceedings of the Seventh SIAM

[2] International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota, USA. SIAM 2007.

[3] Adamek, T. and O'Connor, N.E. A multiscale representation method for nonrigid shapes with a single closed contour. IEEE Circuits and Systems for Video Technology,2004, 14(5): 742-753.

[4] Eamonn J. Keogh, Li Wei, Xiaopeng Xi, Sang-Hee Lee, Michail Vlachos: LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures[C]. In Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006. VLDB 2006: 882-893

[5] Zhang D, Lu G. Review of shape representation and description techniques. Pattern Recognition, 2004,37(1):1-19

[6] Jeremy Buhler, Martin Tompa. Finding motifs using random projections[C]. In Proceedings of the Fifth Annual International Conference on Computational Biology, April 22-25, 2001, Montreal, Canada. ACM, 2001, 67-74.

[7] Lin J., Keogh E., Lonardi, S,et al. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms[C]. In Proceedings of the 8th ACMSIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003, 2-11.

[8] Keogh E., Chakrabarti K., Pazzani M, Mehrotra S. Locally adaptive dimensionality reduction for indexing large time series databases[C]. In proceedings of ACMSIGMOD Conference on Management of Data. Santa Barbara, CA, May 21-24, 2001, 151-162.

[9] http://www.efg2.com/Lab/Graphics/PolygonArea.htm, June 2001.

[10] Xiaopeng Xi, Eamonn Keogh, Li Wei, Agenor Mafra-Neto. Finding Motifs in Database of Shapes, 2007.

Published by Atlantis Press, Paris, France.