Engineering Letters, 14:1, EL_14_1_30 (Advance online publication: 12 February 2007)

Multiresolution Knowledge Mining using Wavelet Transform

R.Pradeep Kumar, P.Nagabhushan

***Abstract*— Most research in Knowledge Mining deal with the basic models like clustering, classification, regression, association rule mining and so on. In the process of quest for knowledge most of the knowledge mining algorithms end up in generating global knowledge while losing focus on the local knowledge. This happens oftenly due to two reasons. First reason is due to dimensionality reduction. The problem of dimensionality reduction has been viewed as the reduction of features to the maximum extent possible while being able to retain the information conveyed by the data set. But most of the dimensionality reduction techniques reduce the dimensions keeping only the retention of global knowledge in mind while compromising with the loss of local knowledge. Second reason is due to optimized feature selection for making a global classification while not being bothered about the intra class relationship. In this paper we present methodologies using wavelet transform for overcoming the loss of local knowledge along the process of mining. First we propose a discrete wavelet transform based multi resolution approach to capture the local knowledge along the process of dimensionality reduction and also being capable of representing the global knowledge with few number of wavelet coefficients. A comparison with PCA has also been made here, which strongly supports our technique using discrete wavelet transform to produce more number of local knowledge and also no/minimum misleading information. Secondly we propose continuous wavelet transform based multiresolution approach for knowledge mining through a novel histogram distance measure. Along both this approach we also propose methodologies for capturing knowledge packets from different sources and integrating this knowledge for generating a comprehensive knowledge base. In the first part knowledge packets are generated along individual sources, filtered and then integrated. In the second part, summing the distances between the objects when observed from different sources does the integration and then the knowledge is mined from overall distance matrix to obtain comprehensive knowledgebase. These foundational techniques are illustrated with a set of '8-O-X' spatial dataset.**

***Index Terms*— Data Mining, Knowledge Mining, Multiresolution Knowledge, Stability factor, Wavelet Transform**

Manuscript received April 14, 2006.

R.Pradeep Kumar is with Department of Studies in Computer Science, University of Mysore, Mysore, India – 570006.(e-mail: rpradeep25@lycos.com).

P.Nagabhushan is with Department of Studies in Computer Science, University of Mysore, Mysore, India – 570006.(e-mail: pnagabhushan@hotmail.com).

I. Introduction

The curse of dimensionality refers to the exponential growth of hyper volumes as a function of dimensionality. All problems become harder as the dimensionality increases. In other words "more the amount of information or details that exist about the samples poorer will be the performance of the classifier". So it is always required to extract optimum amount of details that is necessary for mining or extracting the knowledge. A straightforward method like principal component analysis is commonly used for dimensionality reduction. Most research on dimensionality reduction deal with it only from the point of memory reduction while being able to preserve the global knowledge. Extraction of global knowledge may be appreciable by the non-critical pattern recognition applications. But when it comes to critical pattern recognition applications like signature verification in a bank cheque, the extraction of local knowledge may become highly crucial. Critical PR applications would need the knowledge extracted from both the finer details and also from the coarser details. The finer details would express the local knowledge and the coarser details would express the global knowledge. Thus dimensionality reduction should not be achieved by totally turning a blind eye towards the higher dimensions and an open eye only to the knowledge extracted in the lower dimensions. Dimensionality reduction has to be a process through which we are able to extract the knowledge all along from the higher dimensions to the lower dimensions rather than reduction of dimensions to the maximum extent and looking for some left out global knowledge.

In the first part of the paper we propose a methodology for extracting local and global knowledge along the process of dimensionality reduction [12] using discrete wavelet transform. As Wavelet transform [1-5, 8-11] has been successful in signal processing application a continuity relationship connecting (i-1), i and (i+1) feature sets is preferred, as it exists in case of signals. Such a signal can be usually obtained in case of spatial data. The signal obtained by sampling the spatial data along a particular direction is called as a spatial signal in our context. The localization property of wavelet transform has the capability of extracting the finer details from the spatial signals. These finer detail features are processed and approximated thus extracting the knowledge at all levels of wavelet decomposition of the spatial signal. We also achieve dimensionality reduction by representing the details with few number of wavelet coefficients by preserving the global knowledge. We propose a two stage strategy for knowledge mining (1) Extraction of knowledge packets during the sequence of dimensionality reduction using wavelet transform (2) Fusion of packets from multiple sources for comprehensive knowledge mining. In our experimentation the spatial signal is obtained by sampling the object of interest along all the four direction as illustrated in section 2. It is viewed as four signals being recorded by four sensors or four signals obtained from four different sources. In section 3 we discuss how to extract the knowledge packets from the clusters obtained by a linkage algorithm applied on the data set generated at different wavelet decomposition level. Then we discuss how these interesting knowledge packets obtained along the four directions or from four sources can be fused to generate comprehensive knowledge base, which gives valid, novel, and interesting information.

Knowledge mining [15-25] refers to the overall process of extracting high-level knowledge from low-level data in the context of huge multi dimensional databases. In the process of quest for some knowledge most of the knowledge mining algorithms ends up in generating global knowledge while losing focus on the local knowledge. These local knowledges may be very important in critical applications like medicine, signature verification, defense and so on. By the phrase 'local knowledge' we like to emphasize that the finer information of a sample/object/record in a dataset/database can bring in close association between a set of objects, which may turn to be very interesting. Figure 1 gives a better illustration of our statement on local knowledge.

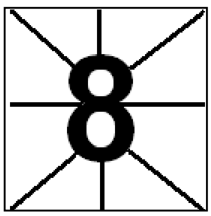


Fig1(a)

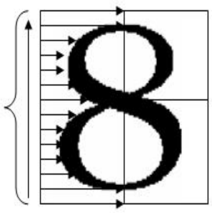


Fig1(b)

In figure 1(a) we have extracted only 8 features whereas in figure 1(b) we collect the features along the horizontal axis moving from bottom to top along the vertical axis obtaining one set of 200 features for a 200 x 200 image. Likewise the features along the four directions (left, top, right and bottom) can be extracted resulting in around 800 features around the object. Bellman Ford's theory on 'Curse of Dimensionality' emphasizes that all problems become harder as the dimensionality increases. So most of the feature extraction techniques always choose a set of optimal features like the one shown in figure 1(a) to classify the samples. But in this process of choosing an optimal set of features we lose the finer information about the object and thus end up obtaining only the global knowledge i.e. we would be able to infer that this particular sample shown in figure 1(a) is an '8' and nothing beyond that. Consider a simple problem of generating association between a set of samples representing numeral '8' and '0' but of different fonts. With the kind of features extracted as shown in figure 1(a) we would be able to classify the samples into two groups with one cluster containing all '8's and another cluster containing all '0'. But when the question regarding the association of these samples within a cluster arises, we lack finer information about the fonts in the feature set which restricts us from inferring further/local knowledge in the sample set. For this, the feature extraction of the kind shown in figure 1(b) is required where the feature set carries information about the entire object. But processing the whole set of features (nearly 800 features) would be a non-trivial process. This is where a procedure to process this entire feature set by keeping in mind the available computational power is required. It is here, the second part of this paper would make a novel proposal to perform knowledge mining by keeping the finer information of the samples intact throughout the mining process. The results obtained out of this proposed method shows that apart from classifying the samples depending on whether it is an '8' or '0' it also maintains the intra cluster associations (similarities between fonts) intact. This is made possible by using continuous wavelet transforms and representing the information conveyed by the continuous wavelet coefficients as symbolic objects (in this case Histograms). The features that are extracted as shown in figure 1(b) when plotted along a spatial axis results in a spatial signal. Through continuous wavelet transform we extract the finer and approximation details at different scales. The localization property of wavelet transform has the capability of extracting the finer details from the spatial signals. In section 4 we discuss how histograms are generated from continuous wavelet transform coefficients and subsequently propose a histogram distance measure in section 4.2 for measuring the distance between the objects along a particular direction. Likewise the distances between the objects from all the directions are added to obtain the overall distance. Then knowledge is mined from these overall distance matrices obtained at different scales. It is also shown how the intra cluster similarity (in this case similarity between fonts that is considered for illustration) is maintained in the comprehensive knowledge base.

II. SPATIAL SIGNAL AND FEATURE EXTRACTION

A set of spatial data comprising the characters 8, O, X has been considered for experimentation purpose. It is because of the overlapping nature of the data with respect to symmetry and the overlapping characteristics due to curvature similarity between '8' and 'O' and linear similarity (crossing lines) between '8' and 'X'. This indicates the high mix up in the samples that have been considered and it has always been a challenge to deal with the '8OX'dataset right from the time an '8OX' dataset was introduced by Anil Jain. The spatial data is of 200 x 200 in size. The samples considered are shown in figure 2 along with the font label.

From these samples four spatial signals were generated by sampling at 200 locations along each of the four directions. In other words 200 feature points along all four directions were recorded. Thus the data set consists of a total of 4 * 200 features. As illustrated in the next section the knowledge is extracted along each direction operating on each of the 200 features individually and later fusing them to generate knowledge about the object as a whole. The features extracted are illustrated in figure 3.

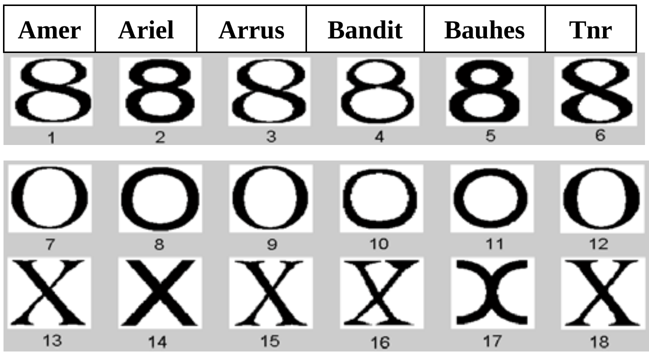
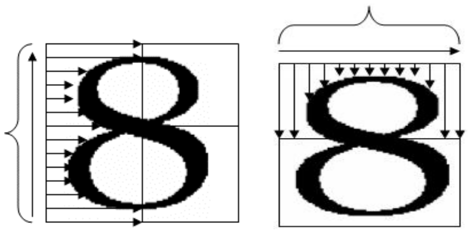


Fig 2: Sample data set with font names and the index



a b

Fig 3: 200 features (***the number of white pixel before the first black pixel***) (a) along the horizontal direction from bottom to top were extracted. (b) along the vertical direction from top to bottom

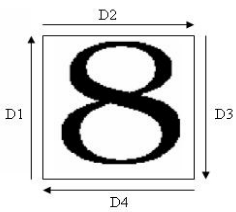


Fig 4: D1, D2, D3, D4 shows the direction along which the spatial object is sampled

The plot of the feature points of the 'character- 8 font-amer' along the four directions is given in figure 5. As shown in figure 4 four spatial signals are obtained for each of the sample with each spatial signal formed out of 200 features recorded along the directions D1, D2, D3 and D4. Now each set of features are subjected to discrete wavelet transform and the knowledge is extracted at different levels of decomposition based on the stability of the knowledge.

III. Knowledge Mining using Discrete Wavelet Transform

Knowledge mining [15–25] refers to the overall process of extracting high-level knowledge from low-level data in the context of huge multi dimensional databases. In signal processing fields, people usually thought wavelets to be convolution filters that have some special properties such as quadrature mirror filters (QMF) and high pass filter. It is agreed that it is convenient to apply wavelets to practical applications if we thought wavelets to be convolution filters. But according to our interpretation from knowledge mining point we consider wavelets as a function which owns some special properties such as compact support, vanishing moments, multi-scaling etc. Compact support guarantees the localization of wavelets, which extracts the finest of information at lower scales; vanishing moment guarantees wavelet processing could distinguish the essential information from non-essential information. It is these features along with hierarchical representation and manipulation, along with feature selection makes wavelet transform an important component in knowledge mining and dimensionality reduction.

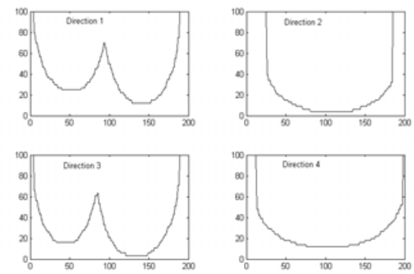


Fig 5 : Spatial signals recorded by sampling along four directions of spatial data of '8'

Approximation vectors and Detailed Vectors in Wavelet transform

Consider the set of all N dimensional, real valued vectors of the form $x = [x_1, x_2, \ldots x_N]$. This set forms an N-dimensional linear vector space and there exist N linearly independent basis vectors a1, a2, ….,aN such that any vector in this space can be expressed as a linear combination of these basis vectors. In other words, there exist unique scalars $\alpha_1, \alpha_2, \ldots \alpha_N$. such that $x=\alpha_1 a_1+\alpha_2 a_2+\ldots+\alpha_N a_N$. Let us call this N-dimensional vector space V_N. We next look at approximating vectors in N-dimensional space by vectors in a subspace of a lower dimension. Suppose we generate all linear combination of just N-1 of the N basis vectors, say $a_1, a_2, \ldots a_{N-1}$. The set of all such linear combinations is also a vector space. It is of N-1 dimension and every vector in it also lies in V_N. It is thus a vector subspace of V_N. Let us call this V_{N-1}. By continuing to drop the last basis vector at each step, we can similarly construct subspaces $V_{N-2}, \ldots V_1$ of dimension N-2,…..,1 respectively. The subspaces V_1 has as its basis the single vector a1. These vector spaces form a nested sequence of subspaces, $V_1 \subset V_2 \subset \ldots \subset V_N$.

Suppose we want to approximate a vector x in V_N by a vector in V_{N-1}. The best possible approximation is the minimum mean squared error or least squares sense is made by choosing that vector (say ζ_{N-1}) in V_{N-1} for which the length of the error vector $e_{N-1} \equiv x - \zeta_{N-1}$ is minimized.

Thus $\zeta_{N-1},\zeta_{N-2}.......\zeta_1$ are the approximation of the vector x at different levels and e_{N-1}, $e_{N-2},......,e_1$ can be seen as the amount of detail that is lost in x in going to its approximation ζ_{N-1} because $x-e_{N-1} = \zeta_{N-1}$. Thus the original vector x can be obtained through the following equation

$$x = e_{N-1} + e_{N-2} + \quad +e_1 + \zeta_1$$

where ζ_1 is the coarsest of approximation.

In wavelet transform the approximation from N dimensions to the lower dimension generates the approximation coefficients and the error vectors generate the corresponding detail coefficients. Thus the entire process of multi resolution knowledge mining involves analyzing the error vectors at different vector spaces V_k where $k < N$ for being capable of representing a stable knowledge. Stable knowledge (a cluster of a set of objects) is one which gets generated and remains undisturbed / uninfluenced by other objects in the data space for sufficient period. In the process of dimensionality reduction we search for the appropriate error vector e_k in the lowest K dimensional space where $K < N$ that would hold the optimum or enough detail for classifier and also extract the knowledge expressed by the error vectors at different dimensions.

3.1 Multiresolution Knowledge Miner

The motivation of Multiresolution analysis is to use a sequence of embedded subspaces to approximate $L^2(R)$ so that people can choose a proper subspace for a specific application task to get a balance between accuracy and efficiency. Higher subspaces can contribute accuracy or finer details but waste computing resources on the other hand lower subspaces would provide approximate or global knowledge at lesser computational cost. Wavelet is related to MRA because of the scaling function ϕ which easily generates a sequence which can provide a simple multiresolution analysis. A direct application of multiresolution analysis is the fast discrete wavelet transform algorithm called the pyramidal algorithm. The idea is to progressively smooth the data using an iterative procedure and keep the detail along the way. After the first decomposition, the data are divided into two parts: one is of average information (projection in the scaling space V_2) and the other is of detail information (projection in the wavelet space W_2). We then repeat the similar decomposition on the data in V_2, and get the projection data in V_1 and W_1, etc. The detail information otherwise termed as the surprise information is used as the feature values for mining the knowledge. In our method we mine the knowledge from the finest detail level to the coarsest detail level. Before we get on to the algorithm we would like to define the term *Stability factor*.

Stability factor:
The stability factor of a node introduced in the above algorithm is defined as the difference between the length of span taken for a group of samples to cluster and the length of span it remains stable without allowing any other samples to join the cluster. Consider the following simple dendrogram

Stabiltiy factor (Sf) of the group {a1, a2, a3} $\}$ = ls – lt

where lt is the length of span/stretch taken for a group to cluster and ls is the stretch/length of span for which the group is stable without allowing any other sample to join the cluster. if Sf > 0 then cluster {a1, a2, a3} is a stable knowledge else cluster {a1, a2, a3} is an unstable knowledge

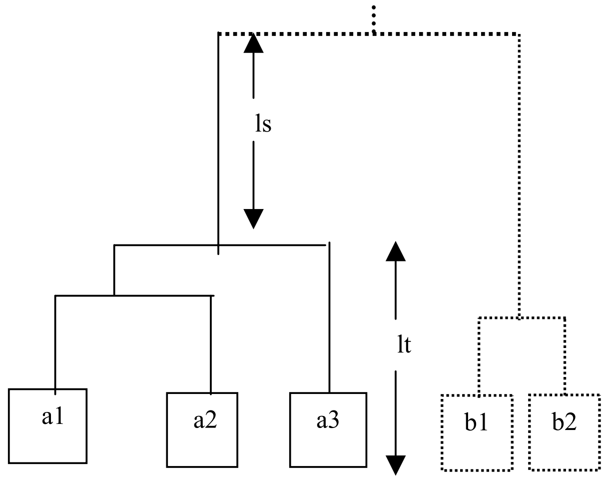


Fig 6: A simple dendrogram

The algorithm for generating knowledge packet is given below

Step 1: Extract the spatial signals S1, S2, S3, and S4 along all the four directions of the characters.

Step 2: For each of the signal Sx for x = 1 to 4

Apply discrete wavelet transform.

Step3: For each decomposition level from 1 to maximum level

Considering the detail wavelet coefficients as the features apply an agglomerative clustering algorithm – linkage algorithm

Step 4: Calculate the ***stability factor*** for all the nodes except the leaf nodes in the dendrogram generated on applying the linkage algorithm.

Step 5: If the ***stability factor*** is greater than 0 label the group of samples or cluster formed under the node as an interesting node else it is a non-interesting node.

Step 6: Fuse the knowledge generated along the four directions to obtain the comprehensive knowledge about the sample set.

The stability factor of a knowledge packet is also the measure of interestingness of the knowledge packet. The interesting measure is denoted by Int_M. Few of the dendrograms obtained after applying linkage clustering for the detail coefficients obtained from wavelet transform of the spatial signal are shown below. The spatial signal obtained along direction 1 and direction 2 are considered for mining the knowledge about ‘8OX’ dataset.

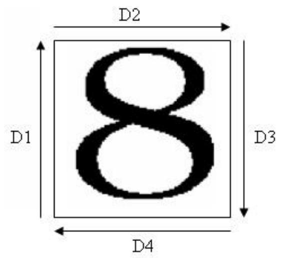


Fig 7 : D1, D2, D3, D4 represent the four directions along which the spatial signals were recorded

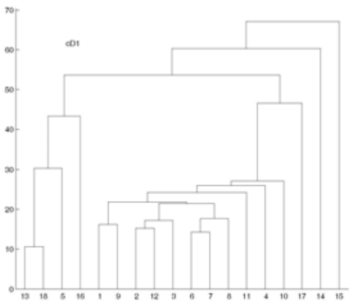


Fig 8(a): Cluster formations with the first level decomposition detailed coefficients

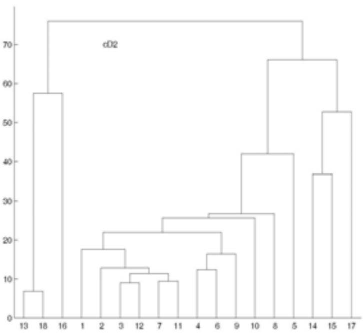


Fig 8(b): Cluster formations with the second level decomposition detailed coefficients

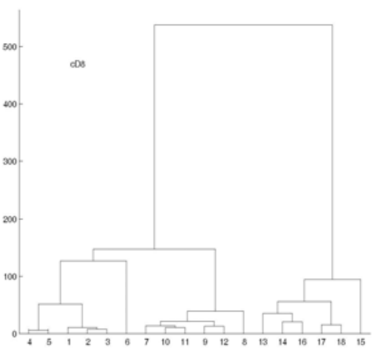


Fig 8(c): Cluster formations with the eighth level decomposition detailed coefficients

Table 1

Knowledge generated from spatial signals recorded along Direction I

Set No	Knowledge Packets	Int_M	Level in which Knowledge found
K1	13, 18	9.19	1
K2	13, 18	43.83	2
K3	13, 18	42.15	3
K4	13, 18	35.81	4
K5	10, 11, 12	5.73	4
K6	2, 4	5.63	4
K7	10, 11, 12	0.44	5
K8	13, 14, 15, 16, 18	47.49	5
K9	7,8,9,10,11,12	64.00	6
K10	1, 2, 3, 4, 5, 6	23.19	6
K11	13, 14, 15, 16, 17, 18	28.75	6
K12	8, 10	18.40	7
K13	11, 12	0.52	7
K14	16, 18	5.91	7
K15	7, 8, 9, 10, 11, 12	10.48	7
K16	6, 13, 14, 15, 16, 17, 18	3.14	7
K17	4, 5	40.1	8
K18	1, 2, 3	29.82	8
K19	17, 18	25.81	8
K20	7, 8, 9, 10, 11, 12	69.97	8
K21	1, 2, 3, 4, 5	23.91	8
K22	13, 14, 15, 16, 17, 18	348.59	8
K23	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	242.60	8

3.2 Fusion of Knowledge from different sources

The knowledge obtained from the four directions are integrated by a set of procedures which involves tree building, node shedding, node collection and then Cross intersection. The following algorithm is executed for integrating the interesting details

Binary tree building for the set of knowledge obtained along a direction

Step 1: Create a root node with all the samples present in the node initially. Let the set of elements be denoted as T.

Step 2: Sort the knowledge obtained based on interesting measure

Step 3: Create a left child node containing the elements with the knowledge of highest interesting measure. Assume the set of elements in left child node as A.

Step 4: Create a right child node with the remaining elements i.e. with the elements T – A

Step 5: For each of the leaf nodes from left to right follow the steps 2- 4 by considering the knowledge one after the other. If the subset of the leaf nodes cannot be created for any knowledge then stop the algorithm.

Table 2

Knowledge generated from spatial signals recorded along Direction II

Set No	Knowledge Packets	Int_M	Level in which Knowledge found
K1	7, 8, 9, 10, 11, 12	1.30	2
K2	10, 11	19.41	4
K3	7, 9, 12	0.77	4
K4	4, 5	29.39	4
K5	10, 11	7.04	5
K6	7, 9, 12	3.69	5
K7	4, 5	79.50	5
K8	7, 8, 9, 10, 11, 12	73.77	5
K9	1, 3, 6	0.51	5
K10	4, 5	7.14	6
K11	2, 4, 5	17.43	6
K12	1, 3	72.68	7
K13	4, 5	134.42	7
K14	6, 7, 8, 9, 12	0.41	7
K15	13, 14, 15, 16, 17, 18	49.42	7
K16	4, 5	3.19	8
K17	8, 10, 11, 12	1.21	8
1K8	4, 5, 8, 10, 11, 12	5.63	8
K19	13, 14	45.26	8
K20	15, 18	40.31	8
K21	1, 2, 3, 6, 7, 9	10.58	8
K22	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17	27.18	8

The tree building process is illustrated below with the knowledge obtained along direction 1

Step 1:

Create a node with all the elements.

T = { 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}

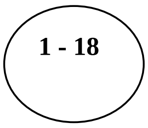


Fig 9(a)

Step 2:

Table 1 is sorted based on their interesting measure.

Step 3:

Create a left child node containing the elements with the knowledge of highest interesting measure. As given in table 1 the knowledge with highest interesting measure is with the set of elements 13-18 with an interesting measure of 348.59.

A = {13,14,15.16,17,18}

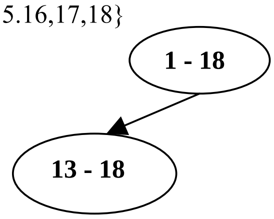


Fig 9(b)

Step 4:

Create a right child node with the remaining elements i.e. with the elements T – A. Now the right child will consist of

T - A = { 1,2,3,4,5,6,7, 8,9,10,11,12}

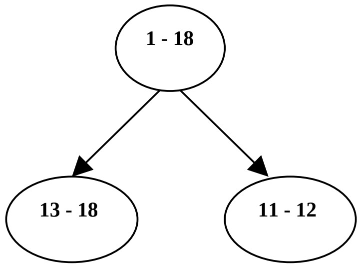


Fig 9(c)

Step 5:

For each of the leaf node repeat the steps 2-4 by considering the knowledge in ascending order. If the entire objects in a node comprise of knowledge then the left child of the node would carry the entire set of objects while the right child would carry a null set. The final tree obtained looks like as shown in figure 10. Nodes B, B1and B2 of figure 10 illustrate the case where node B1 comprise the entire set of objects in B which is a knowledge and node B2 is a null set.

Knowledge filtering:

The Knowledge filtering phase involves node shedding and node collection. It can be observed that all the nodes forming the left child are knowledge obtained along a direction and the right nodes consist of the remaining elements. So we shed the right nodes and collect the left nodes. Thus the filtered knowledge obtained along direction-I is as given below

KD-I = {{13, 14, 15, 16, 17, 18}, {13, 14, 15, 16, 18}, {7, 8, 9, 10, 11, 12}, {13, 18}, {8, 10}, {4, 5}, {11, 12}, {1, 2, 3}};

Likewise the filtered knowledge obtained along direction-II is

KD-II = {{4, 5}, {7, 8, 9, 10, 11, 12}, {10, 11}, {1,3}, {7, 9, 12}, {13, 14, 15, 16, 17, 18}, {13, 14}, {15, 18}};

T = { 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}

A= {13, 14, 15, 16, 17, 18}

B= T – A = { 1,2,3,4,5,6,7,8,9,10,11,12}

B1 = {1,2,3,4,5,6,7,8,9,10,11,12}

B2 = B – B1 = { }

C= {13, 14, 15, 16, 18}

D = A – C = {17}

E = {7, 8, 9, 10, 11, 12}

F = B – E = {1, 2, 3, 4, 5, 6}

G = {13, 18}

H = C – G = {14, 15, 16}

I = {8, 10}

J = E – I = {7, 9, 11, 12}

K = {4, 5}

L = E – I = {1, 2, 3, 6}

M = {11, 12}

N = J – M = {7, 9}

O = {1, 2, 3}

P = L – O = {6}

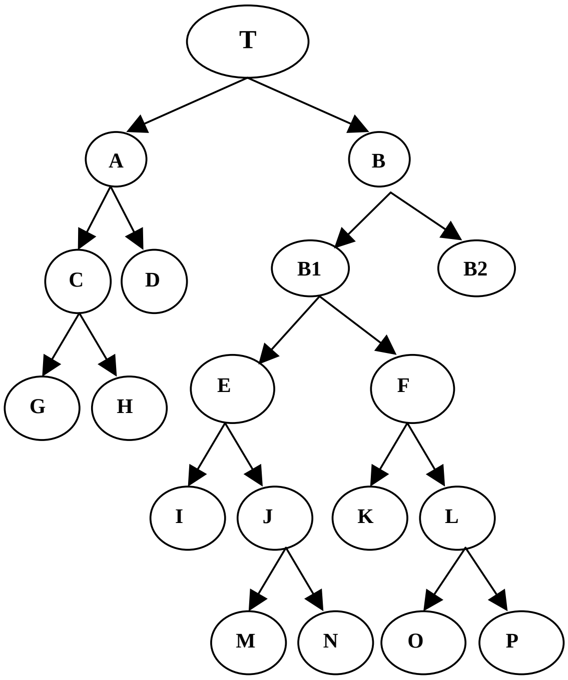


Fig 10: Knowledge Filter Tree

Cross Intersection:

Cross Intersection of two sets A and B represented by CI (A, B) can be defined as follows

Given two sets A and B with a set of sets say A = {{a1}, {a2} . . ., {aN}} and B = {{b1}, {b2} . . ., {bN}} then their cross intersection is given by

CI (A, B) = {{a1∩b1}, {a1∩b2},...,{a1∩bN} ,

{a2∩b1},{a2∩b2},...,{a2∩bN} ,

: : : :

: : : :

: : : :

{aN∩b1},{aN∩b2},... ,{aN∩bN}}

The null sets in CI(A,B) can be eliminated. Thus the integration of the two sets of knowledge KD-I and KD-II are done using cross intersection

Integrated / Comprehensive Knowledge

IKD = CI (KD-I, KD-II)

After the process, the comprehensive stable knowledge obtained is

IKD = {{13, 14, 15, 16, 17, 18}, {7, 8, 9, 10, 11, 12}, {4, 5}, {1, 3}, {10, 11}, {7, 9, 12}, {13, 14}, {15, 18}, {13, 14, 15, 16, 18}, {13, 18}, {8, 10}}

3.3 Discussions

Discussion on IKD:

IKD(1) holds all the samples of character ‘X’.

IKD(2) holds all the samples of character ‘O’.

IKD(3) brings together samples 4 and 5 which is a local knowledge within IKD(1) because the shape of the character ‘8’ of these two samples look very much similar and are dissimilar from the other samples holding the character ‘8’.

Likewise IKD(4) are similar looking ‘8’, IKD(5) are similar looking ‘O’, IKD(6) are similar looking ‘O’ , IKD(8) and IKD(10) are similar looking ‘X’ , IKD(11) are similar looking ‘O’.

IKD(9) is also a very interesting knowledge because it holds all the samples holding the character ‘X’ except sample 17 which is of a different shaped ‘X’.

Apart from IKD there can also be some knowledge extracted from the knowledge tree by collecting the nodes with a single element. These single element nodes give very interesting and surprising information. If the tree drawn with the knowledge extracted along direction I is observed node ‘D’ and node ‘P’ consist of single sample elements 17 and 6 respectively. If the sample 6 is observed carefully it is the only sample with both the upper and the lower portion of ‘8’ of same size whereas in all other samples holding ‘8’ the upper part is slightly smaller than the lower part. Likewise sample 17 is the only ‘X’ with curved shape whereas all other samples holding ‘X’ maintains linearity along its shapes.

The above discussion establishes that the knowledge extracted by using a multi-resolution approach using wavelets are very interesting and are also powerful enough to extract both the global and local knowledge along the process of dimensionality reduction.

Comparison with PCA:

Dimensionality reduction achieved through PCA is an irreversible process while the one we have proposed using wavelet transform can be made reversible. Also PCA is capable of retaining only the global knowledge while loosing all components pertaining to the local knowledge. On achieving dimensionality reduction through PCA and then applying the clustering technique as before similar samples are never found to come together locally within a cluster and some of the local knowledge packets obtained are totally unacceptable due to misleading information generated. The stable knowledge obtained along direction I and direction II obtained from the first few principal component is given below along with the dendrograms shown in figures 11 and 12.

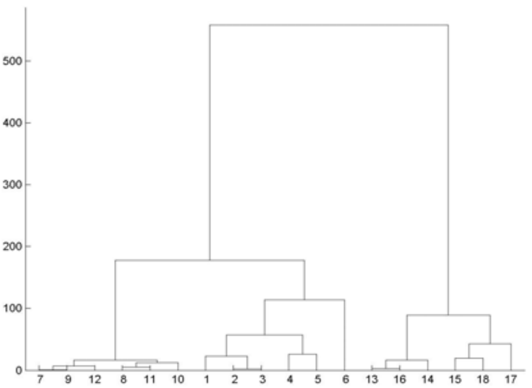


Fig 11: Cluster formations with the first three Principal components obtained from the data set recorded along direction-1

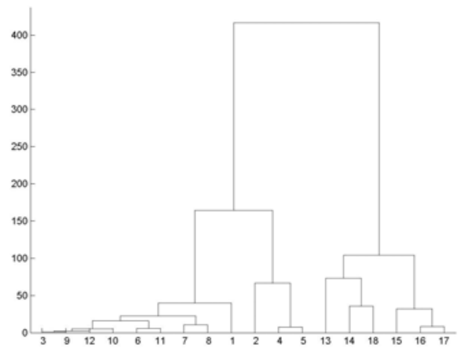


Fig 12: Cluster formations with the first three Principal components obtained from the data set recorded along direction-2

KD-I $_{PCA1}$ = {{13, 14, 15, 16, 17, 18}, {13, 14, 16}, {13, 14}, {7, 8, 9, 10, 11, 12}, {7, 9}, {2, 3}}

KD-II $_{PCA1}$ = {{13, 14, 15, 16, 17, 18}, {15,16,17}, {16,17}, {1, 3, 6, 7, 8, 9, 10, 11, 12}, {3, 9, 12, 10}, {4, 5}}

The knowledge obtained along both the direction are integrated though cross intersection

IKD $_{PCA}$ = CI (KD-I $_{PCA1,}$ KD-II $_{PCA1}$)

= {{13, 14, 15, 16, 17, 18}, {13, 14, 16}, {13, 14}, {7, 8, 9, 10, 11, 12}, {7, 9}, {15, 16, 17}, {4, 5}}

Discussion on IKD $_{PCA}$

IKD $_{PCA}$ (1) is a global knowledge

IKD $_{PCA}$ (2) is not an interesting pattern which can be observed from the shapes of character '8' in samples 13, 14, 16. It is a misleading knowledge

IKD $_{PCA}$ (3) is a local knowledge

IKD $_{PCA}$ (4) is a global knowledge

IKD $_{PCA}$ (5) is a local knowledge

IKD $_{PCA}$ (6) is a misleading knowledge

IKD $_{PCA}$ (7) is a local knowledge

Thus out of the seven knowledge obtained there are 2 global knowledge, 3 local knowledge and 2 misleading information. The local and global knowledge obtained through PCA is a subset of the knowledge in the set IKD obtained through multi-resolution approach. There were 2 misleading knowledge generated by PCA whereas not even a single misleading knowledge was generated though multi-resolution approach. Thus a total of 11 interesting knowledge were obtained through multi-resolution approach in which {4, 5}, {1, 3}, {7, 9, 12}, {15, 18},{13, 14, 15, 16, 18}, {13, 18} are highly interesting.

Discussion : DWT for Multiresolution Knowledge Mining

A technique based on multi-resolution approach using wavelet transform to mine the knowledge existing in a sample set with spatial features was proposed. An illustrative example with 18 samples and 200 feature points along four directions was considered for presentation. The experimentation conducted by considering 48 kinds of font for each of the numerals from 0 to 9 have generated breakthrough results. The results establish the strength of the multi-resolution techniques over PCA's approach. This section emphasizes that dimensionality reduction should not be considered as a process of only reducing the features by retaining the global properties but essential efforts need to be taken to retain the local knowledge present in the data. Thus it is very much necessary to propose methodologies for dimensionality reduction where along the process of dimensionality reduction we collect the local knowledge which otherwise would be lost in the reduced feature set. This attempt is one such effort for capturing and retaining the stable local knowledge along the process of dimensionality reduction. This paper also brings out that all knowledge may not be stable or reliable knowledge. Thus methodologies for generating stable knowledge and new methodologies to integrate/fuse the knowledge obtained from different directions or from multiple sensors can be very good contributions in future.

IV. KNOWLEDGE MINING USING CONTINUOUS WAVELET TRANSFORM

The CWT or continuous–time wavelet transform of f(t) with respect to a wavelet $\psi(t)$ is defined as

$$W(a,b) \equiv \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{|a|}}\psi^{*}\left(\frac{t-b}{a}\right)dt$$

where a and b are real and * denotes complex conjugation. Thus, the wavelet transform is a function of two variables. Both f(t) and ψ(t) belong to $L^2(R)$, the set of square integrable functions, also called the set of energy signals. The variable 'b' represents time shift or translation and 'a' referred to as the 'scale' determines the amount of time scaling or dilation. On a careful analysis of the CWT equation it can be clearly viewed as a correlation equation. The set of squared integrable functions forms a linear vector space under addition and scalar multiplication. This vector space comes with a well defined inner product. We see that the CWT is essentially a collection of inner products of a signal f(t) and the translated and dilated wavelet $\psi_{a,b}$ (t) for all a and b where the values W(a,b) represents the correlation coefficients:

$W(a,b) = < f(t) , \psi_{a,b}(t) >$

CWT has a cross correlation interpretation as well

$W(a,b) = <f(t) , \psi_{a,0}(t-b) > = R_{f,\psi a,0}(b)$

It is these interpretations which make wavelet an astonishing transform having larger orientation towards knowledge mining because details on f(t) are captured at multiple scales/resolutions. Most of the time it is said that CWT domain has a redundant representation of f(t). But redundant from what point of view is more important. CWT representation is redundant only when it comes to reconstruction of f(t) i.e. a part of CWT coefficients are enough for an inverse transform to produce f(t). But from knowledge mining point of view what are available in CWT domain are information rich coefficients which will allow us to extract huge amount of knowledge. In this paper we discuss the extraction of knowledge packets at multiple scales or resolution in CWT domain. On the other hand if the coefficients of CWT (f (t)) at a particular scale are more in number the knowledge mining algorithms can become computationally intensive. So we adopt a methodology of aggregation by representing the set of coefficients at each scale for a particular sample by a histogram. In our experimentation our function f(.) is not time based but spatial based which would be elaborated in the next section. So the procedure is clear as follows

Consider a spatial signal f(s)

Compute CWT(f(s))

Generate Histograms of CWT coefficients at scales 2, 4, 8, 16. 32, 64, 128

Thus for each sample we obtain 7 histograms recorded at the above scales.

Derive knowledge packets by comparing the corresponding histograms of different samples recorded at different scales.

4.1 Histogram Generation and Regression

In the previous section it was shown how four spatial signals were extracted by sampling the object along four directions. Now we start considering the signals one by one. First for the signal generated along direction D1 we apply continuous wavelet transform thus to obtain the CWT coefficients at different scales. These coefficients are generated for all the samples by applying CWT on signal generated along direction D1. As mentioned before these coefficients are information rich as they are obtained at multiple scales. Now we start analyzing the association between these samples at scales 2, 4, 8, 16, 32, 64, 128. The maximum scale is decided based on the length of the signal. The maximum scale is chosen in such a way that 2^n < length of the spatial signal. The reason is the coefficients at the intermediate scales would be redundant (coefficients approximately equal to any one of these scales)as the intermediate scales would result in generating the same knowledge as these scales would do. For a particular scale Sx we collect the CWT coefficients of every sample and generate a histogram corresponding to each of them. This histogram is the frequency distribution of the CWT coefficients binned at 10 centres. This decision for obtaining a histogram with 10 centres helps us in aggregating the entire set of 200 coefficients at a particular scale with just 10 values. As can be observed later in the results this aggregation never results in loss of local associations. The maximum and the minimum coefficient values at a particular scale for a sample is recorded and the maximum and minimum among them are considered as the spread for the histogram generated at that scale. This spread is binned at 10 equiwidth positions. Thus for N number of samples given a direction Dx and a scale Sx we generate N number of Histograms. Thus for four directions we have 4N Histograms for a scale Sx. For eight such scales we get 32N histograms. These histograms are the aggregated "information rich symbolic objects[27-28]" obtained from the information wealthy wavelet coefficients. There may be an argument why all of the wavelet coefficients are considered in aggregated information as only the positive values of wavelet coefficients are going to project the similarity. We should understand that here wavelet is a template and the coefficients generated are the measures of comparison between the spatial signal and the wavelet at a particular scale. Here both the similarity and dissimilarity between the signal and the wavelet template are going to generate information with the template as reference. Now the knowledge is mined by estimating the distances between these set of histograms. The procedure can be summarized as follows

{

For signal Sx where x = 1,2,3,4

{

Apply CWT on Sx

For scale Sc = 2, 4, 8, 16, 32, 64, 128

{

Generate Histograms;

Compute distance between Histograms;

Generate the distance matrix for each direction

Add the four distance matrices obtained along 4 directions at a particular scale

Generate knowledge at a particular scale

}

Obtain comprehensive knowledge

}

4.2 A Novel Histogram Distance Measure

The computation of Histogram distance is not a straightforward approach and most conventional distance measures may fail or perform poorly. In this paper we introduce a regression based Histogram distance measure. A histogram distance measure based on cumulative histogram & regression had been proposed by Nagabhushan and Pande[26].In this paper we simplify the formulation of this distance measure by refining the 4 cases they propose into just 2 cases.

Consider a histogram H with 10 bins as shown in fig 13

H → {b1 b2 b3 b4 b5 b6 b7 b8 b9 10}

where bi is the frequency count of the bin centred at Ci.

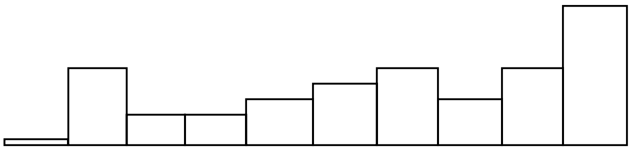


Fig 13 A Sample Histogram H

Now a cumulative frequency distribution is computed for each of these 10 centres resulting in the cumulative histogram

CH → {cn1 cn2 cn3 cn4 cn5 cn6 cn7 cn8 cn9 cn10}

where cni = sum(cnk) for k = 0 to i;

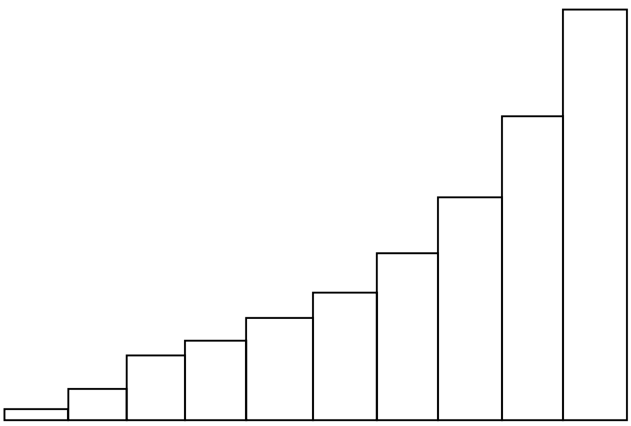


Fig 14 Cumulative Histogram CH

Then the cumulative histogram is normalized by dividing cni for i = 1 to 10 by cn10. Now 10 points are marked on the top of each bin in the cumulative histogram corresponding to the bin centres. Now a 1st order polynomial is fitted across these 10 points to obtain regression lines with yi ranging between 0 and 1 and xi's range is decided by the minimum and maximum coefficient values at a particular scale. It is illustrated in the figure 16.

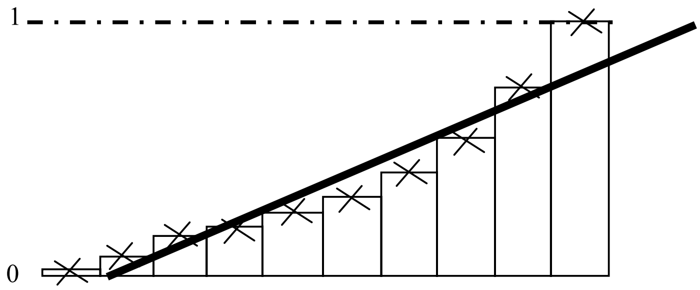


Fig16 Ist order polynomial (Line) fitted across the points with yi ranging between 0 and 1

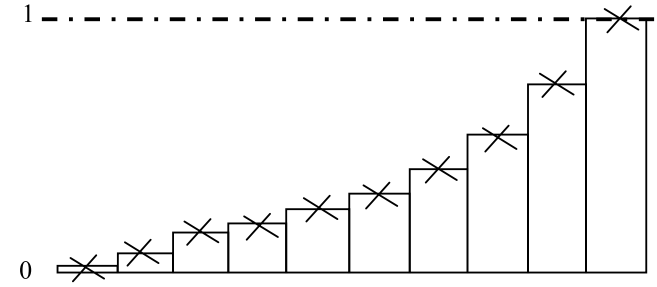


Fig 15 Normalized Cumulative Histogram NCH and the 10 points marked on the top center of each bin

Thus each sample is associated with a regression line for each direction. Now the distance between these samples along a particular direction can be computed by measuring the distance between these lines. For a pair of samples the distance computation would be based on the following two cases

Case 1: Two lines are separated out

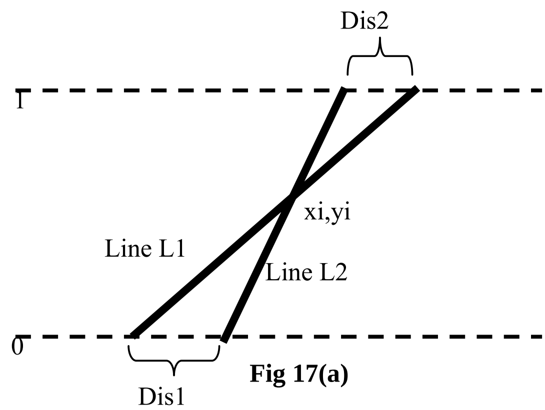


Fig 17(a)

Distance(Line L1, Line L2)

= (Dis1 * yi + Dis2 * (1-yi)) / 2

Case 2: Two lines are intersecting each other

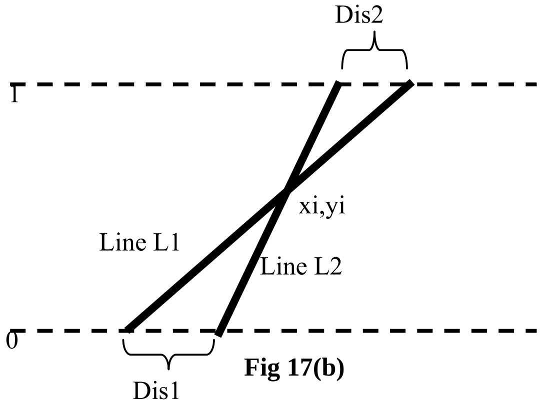


Fig 17(b)

Distance(Line L1, Line L2)

= (Dis1 * yi + Dis2 * (1-yi)) / 2

Likewise the pair wise distances are computed for all four directions and these distances are added to obtain the total distance between two objects. For N number of samples/objects the following procedure is executed to compute the distance

For obj1 = 1 to N-1

{

For obj2 = obj1+1 to N

{

For D = 1 to 4

{

Compute Regression Distance between obj1 & obj 2 along Direction D

}

Total Distance (I,J) = Sum of distance computed along all four directions.

}

}

Thus Total distance is an N x N distance matrix, which will be given as an input to the clustering algorithm for knowledge extraction.

4.3 Discussions

Discussions on comprehensive knowledge

The 18 samples shown in figure 1 were considered for experimentation. Four Regression lines for each of these samples were generated. The pair wise distances between these lines were computed and given as input for the linkage-clustering algorithm. The results produced at various scales/resolutions are given in figure 12. Here too we make use of the stability factor (illustrated in section 3) for mining stable knowledge at every scale.

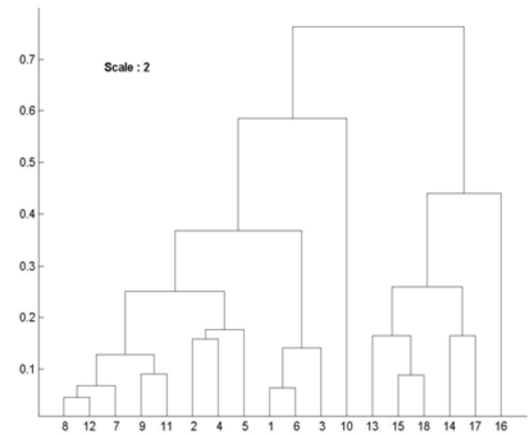


Fig 18 (a) Stable Knowledge at scale = 2: {1,6,3} {7,8,9,11,12}

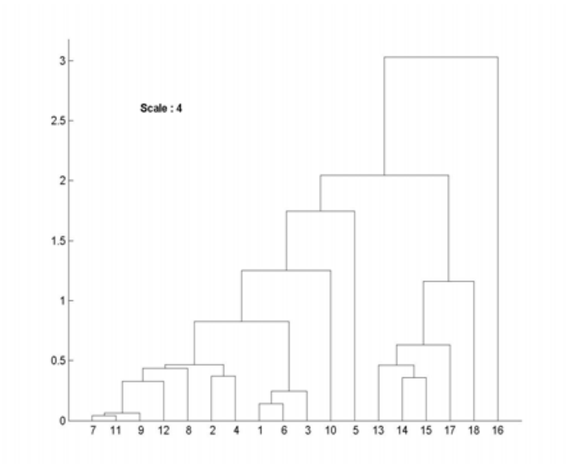


Fig 18 (b) Stable Knowledge at scale = 4: {1,6,3}

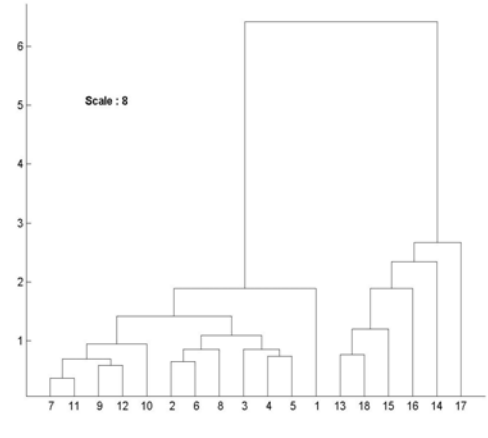


Fig 18 (c) Stable Knowledge at scale = 8:{1,2,3,4,5,6, 7,8,9,10,11,12} {13,14,15,16,17,18}

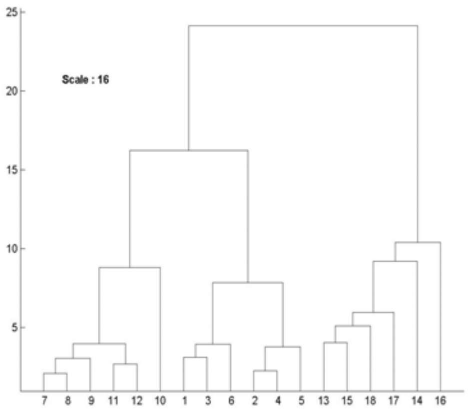


Fig 18 (d) Stable Knowledge at scale = 16: {1,3,6} {2,4,5} {1,2,3,4,5,6} {7,8,9,10,11,12} {13,14,15,16,17,18}

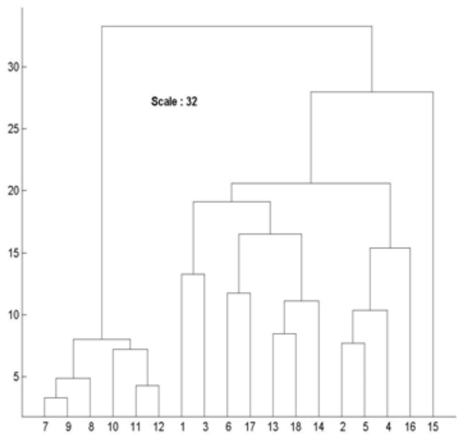


Fig 18 (e) Stable Knowledge at scale = 32: {7,8,9,10,11,12}

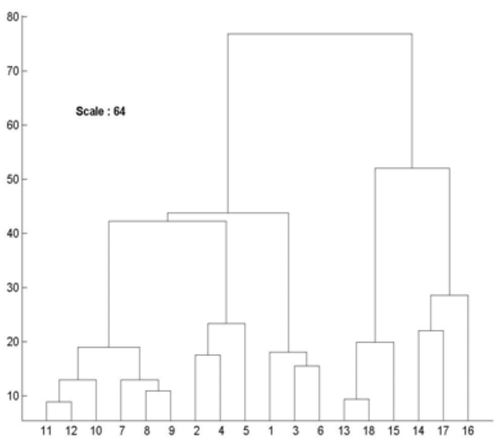


Fig 18 (f) Stable Knowledge at scale = 64: {6,7,8,9,10,11,12} {2,4,5} {1,3,6} {13,18} {13,18,15} {14,16,17}

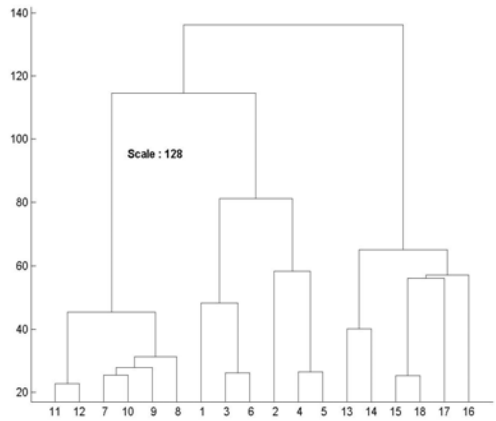


Fig 18 (g) Stable Knowledge at scale = 128: {3,6}{4,5}{15,18} {7,8,9,10,11,12} {13,14,15,16,17,18}

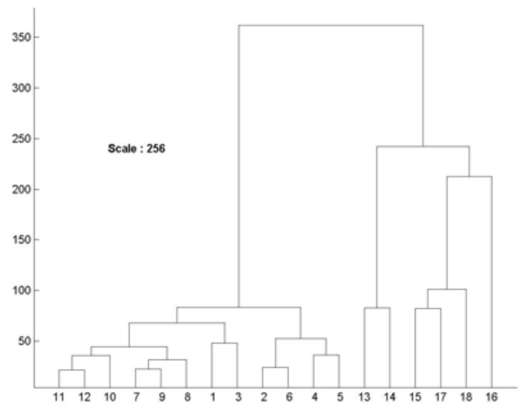


Fig 18 (h) As the length of the signal is 200 we stop mining at a scale n such that n < 200

As can be observed from figure 18 the comprehensive/integrated knowledge would consist of the following sets of stable knowledge

KN1 – {1,3,6}

KN2 – {2,4,5}

KN3 – {1,2,3,4,5,6}

KN4 – {7,8,9,10,11,12}

KN5 – {13,14,15,16,17,18}

KN6 – {13,18}

KN7 – {13,18,15}

KN8 – {14,16,17}

KN9 – {3,6}

KN10 – {4,5}

KN11- {1,2,3,4,5,6,7,8,9,10,11,12}

There are 11 sets of knowledge out of which most of them are interesting and stable. KN1, KN2, KN6, KN7, KN8, KN9, KN10 are all interesting local knowledges or intra cluster groups while KN3, KN4, KN5 & KN11 are global knowledges. KN3 is the group of '8's, KN4 is group of 'O's and KN5 is group of "X's. These results emphasizes that this technique based on continuous wavelet transform and regression based Histogram distance measure mine the global knowledge while keeping the intracluster association intact.

It is obvious when it comes to histogram analysis a few questions regarding flipped objects arises in the mind which raises doubts about this technique where usually the histogram of an object and its flipped version would be same. The wavelet histograms always have the capability to discriminate between flipped objects, as the histograms will also be flipped if the object is flipped. This is the advantage of wavelet histogram compared to conventional histograms. It is illustrated below

Consider a set of sample points

A1 = {1 2 3 4 5 6 7 8 7 6 5 4} and A2 = {4 5 6 7 8 7 6 5 4 3 2 1} which is the flipped version of A1. Now the computation of CWT coefficients for A1 and A2 at scale 2 results in

C1 = {-0.7071 -0.7071 -0.7071 -0.7071 -0.7071 -0.7071 -0.7071 -0.7071 0.7071 0.7071 0.7071 0.7071};

C2 = {-0.7071 -0.7071 -0.7071 -0.7071 0.7071 0.7071 0.7071 0.7071 0.7071 0.7071 0.7071 0.7071}

The count of 0.7071 and –0.7071 are given below for C1 and C2

Table 3

	-0.7071	0.7071
C1	8	4
C2	4	8

which indicates a flip in the object results in flip in histogram. So wavelet histograms will also be useful in identifying the flipped objects.

Discussions: CWT for Multiresolution Knowledge Mining

A technique based on multi-resolution approach using continuous wavelet transform to mine the knowledge existing in a sample set with spatial features was proposed. An illustrative example with 18 samples and 200 feature points along four directions was considered for presentation. The experimentation conducted by considering 48 kinds of font for each of the numerals from 0 to 9 have generated breakthrough results. The results establish the strength of the multi-resolution techniques for mining comprehensive knowledge. The paper emphasizes that consideration of only prominent features for mining application would tend to compromise with local knowledge. This local knowledge would be very important for critical knowledge mining applications. Thus it is very much necessary to propose methodologies for knowledge mining, which would be capable of extracting comprehensive knowledge. Methodologies for refining these comprehensive knowledge for filtering out interesting knowledge can also be future scope of interest. This section is one such effort for capturing and retaining the stable local knowledge along the process of knowledge mining. Another perspective of this methodology is to consider that multiple sensors from 4 different directions record the spatial signals. These spatial signals along each direction can be processed independently in a parallel environment to generate the distances between these samples/objects. Then knowledge can be mined along all directions and fused together as done in the first part of this paper using DWT or the distances can be added and could be given as input to a mining model (in this case clustering model) as done in the second part of this paper.

Discussions: Performance of DWT approach Vs CWT approach for Multi Resolution Knowledge

The knowledge obtained from DWT approach discussed in section 3 and the CWT approach discussed in the section 4 is tabulated blow

Table 4

Knowledge obtained through DWT
{{13, 14, 15, 16, 17, 18}, {7, 8, 9, 10, 11, 12}, {4, 5}, {1, 3}, {10, 11}, {7, 9, 12}, {13, 14}, {15, 18}, {13, 14, 15, 16, 18}, {13, 18}, {8, 10}
Knowledge obtained through CWT
{{1,3,6}, {2,4,5}, {1,2,3,4,5,6}, {7,8,9,10,11,12}, {13,14,15,16,17,18}, {13,18}, {13,18,15}, {14,16,17}, {3,6}, {4,5}, {1,2,3,4,5,6,7,8,9,10,11,12}

It is very interesting to note that all of the local associations/intra-cluster associations among the samples are extracted in the above two approaches. The knowledge packets obtained from both the approaches are non-misleading.

Computational Complexity of DWT Technique is computed as below:

Let Number of sampled points in the spatial signal be = N. DWT of N sample points is order of O(N) complexity. Let Number of objects/samples be = K. Number of sources/directions from which the signals are generated/recorded = D. Therefore the complexity of DWT of D * K signals of N sample points each = O(N * K * D). Linkage algorithm with K number of objects will be of $O(K^2)$ complexity. For D number of sources/directions it is $O(D*K^2)$. Let the number of knowledge packets generated be P. The worst case complexity for Knowledge filtering would be O(P). For D directions it is O(P*D). Then for cross intersection the worst case complexity is $O(P^2)$. Therefore the overall complexity of DWT technique for multiresolution knowledge mining is

$O(N*K*D)+O(D*K^2) + O(P*D) + O(P^2)$.

Computational Complexity of CWT Technique is computed as below:

Let Number of sampled points in the spatial signal be = N. CWT of N sample points for S number of scales for K number of object/samples for D directions is order of O(N*S*K*D) complexity. Histogram generation for each of these scales for K number of objects for D direction is O(N * S * D * K). Regression for b number of bins would take O(B) complexity. Distance matrix computation is of $O(K^2 * S * D)$. Clustering algorithm would consume $O(K^2)$ complexity. Thus the overall complexity of the technique is

$$O(N * S * K * D) + O(K^2*S*D)$$

Experimental Analysis on Fish Contour

The above foundational techniques were tested on the dataset obtained from SQUID database maintained by Mokhtar. This database comprises of fish contours and can be downloaded. The techniques were tested on these fish contours dataset where we followed the same approach as before by obtaining the spatial signals along the four directions and mining at different resolutions. A sample set of the experimental results has been presented here. The results obtained are very impressive and it can be observed that the association between the samples/objects pops up at different resolutions. It leads us to conclude that knowledge can be mined only by looking at the data samples at multiple resolutions. Trying to mine the knowledge at one particular resolution disallows us from capturing many hidded relationships among the samples. Depending on the criticality of the knowledge mining application different resolutions will have to be adopted. The results are presented in figure 19 and 20.Table 4 & Table 5 gives the comprehensive knowledge packets mined.

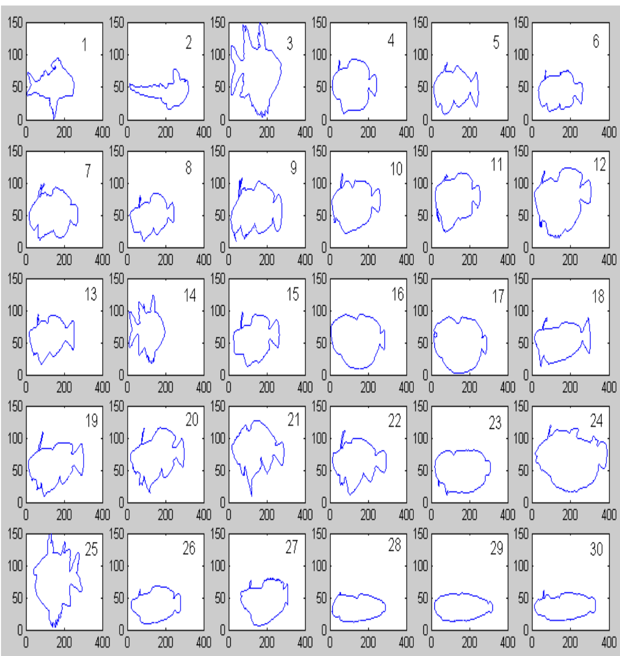


Fig 19 : Fish Contour Samples along with index

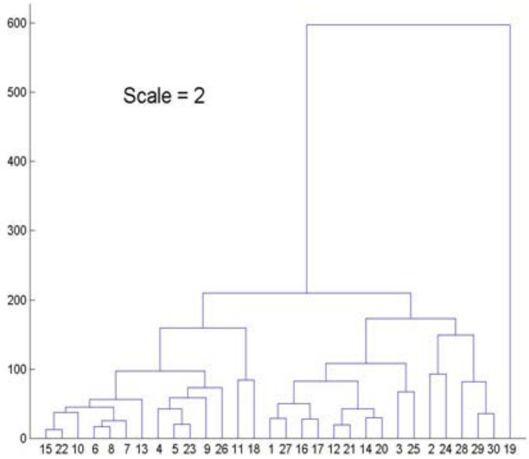


Fig 20(a) Knowledge packets at Scale = 2: {15,22} {5,23} {12,21} {29,30}

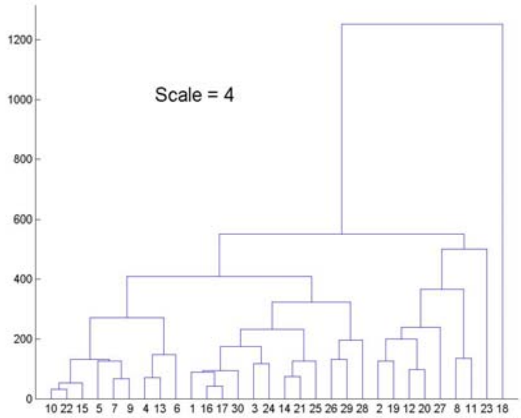


Fig 20(b) Knowledge packets at Scale = 4: {10,22,15} {16,17} {8,11}

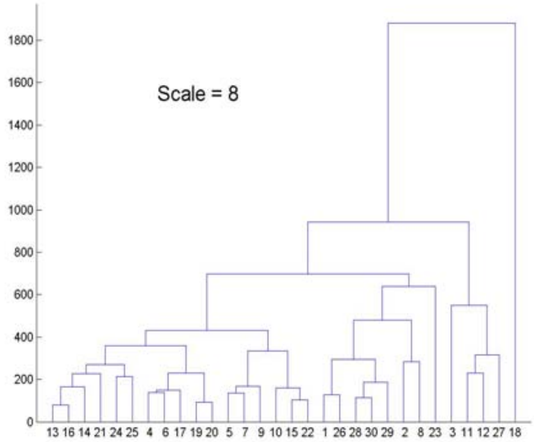


Fig 20(c) Knowledge packets at Scale = 8: {19,20} {1,26} {10,15,22}

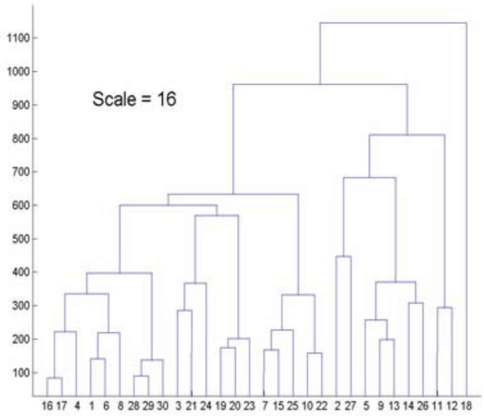


Fig 20(d) Knowledge packets at Scale = 16: {16,17} {28,29,30} {19,20,23} {10,22} {11,12}

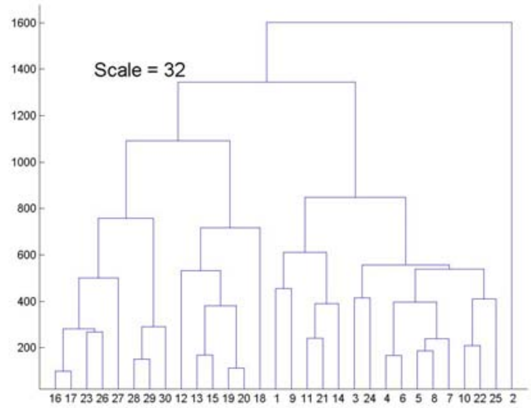


Fig 20(e) Knowledge packets at Scale = 32: {16,17} {28,29,30} {13,15} {19,20} {4,6} {10,22}

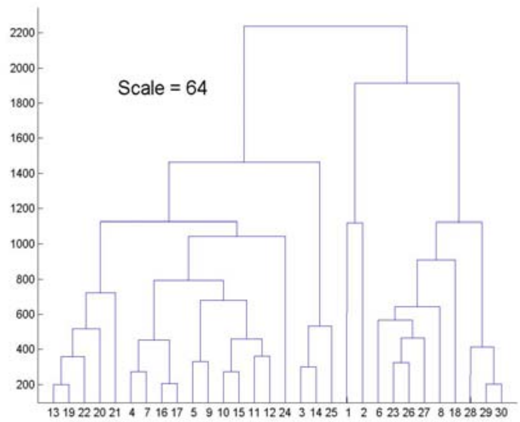


Fig 20(f) Knowledge packets at Scale = 64: {13,19} {16,17} {5,9} {3,14} {28,29,30}

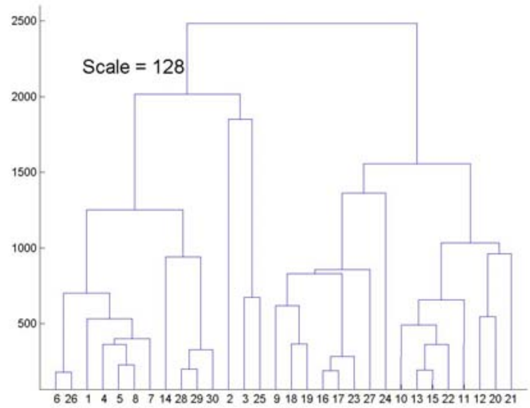
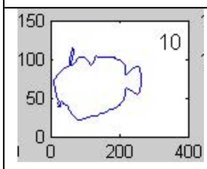
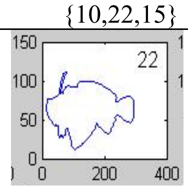
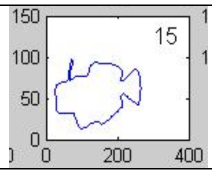
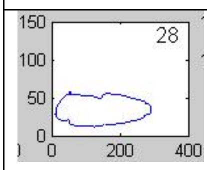
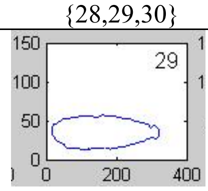
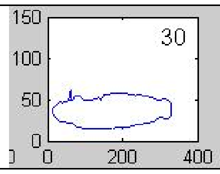
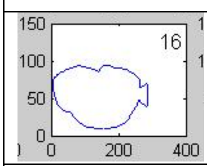
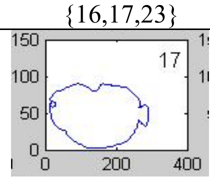
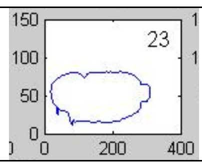
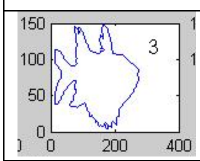
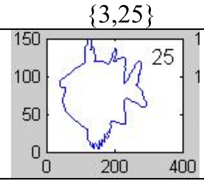


Fig 20(g) Knowledge packets at Scale = 128: {6,26} {28,29,30} {3,25} {16,17,23} {13,15}

The comprehensive knowledge includes the following set of rich local knowledge that are extracted from the 30 input samples that of fish contours through multi -resolution mining

{15,22} {5,23} {12,21} {29,30} {10,22,15} {16,17} {8,11} {19,20} {1,26} {28,29,30} {19,20,23} {10,22} {13,15} {4,6} {13,19} {5,9} {3,14} {6,26} {3,25} {16,17,23}

A few impressive knowledge packets are given in table 5. Out of 20 knowledge packet 19 packets are impressive extracts.

{10,22,15}		
		
{28,29,30}		
		
{16,17,23}		
		
{3,25}		
		

V. Conclusion and Future Work

In this paper two techniques for multiresolution knowledge mining based on Discrete Wavelet Transform and Continuous Wavelet Transform is presented. Many novel ideas have been introduced in terms of knowledge filtering, knowledge integration through cross intersection, aggregation of details of a sample with wavelet histograms, histogram distance measure based on cumulative histogram and regression and so on. It brings out an approach to deal with distributed source or sensors recording data independently and how these data can be mined independently for obtaining knowledge packets and later integrated to obtain the comprehensive knowledge base. We also have emphasized the need for techniques where the intra cluster association between the samples or objects is retained even after dimensionality reduction or feature selection. The results in table 5 indicate that knowledge packets pop up at multiple resolutions. If the data is observed at a single resolution the knowledge packets obtained will be only a fraction of knowledge that would have been embedded in the database. In this paper we have dealt with spatial signals, which can be directly adopted for time-based signals. We have also introduced distance measures, which can be directly adopted for symbolic objects like histogram. Future work in this direction can be proposing techniques for mining multivariate data at multiple resolutions, methods for handling distributed data and knowledge integration etc. Techniques for extracting details from the data at multiple resolutions can also find lot of scope in future.

References

[1] F. Abramovich, T. Bailey, and T. Sapatinas.Wavelet analysis and its statistical applications. *JRSSD*, (48):1–30, 2000.

[2] A. Antoniadis and G. Oppenhiem, editors. *Wavelets and Statistics, Lecture Notes in Statistics*. Springer-Verlag, 1995.

[3] Y. Meyer. *Wavlets–Algorithms and Application*. SIAM, 1993.

[174] R. Young. *Wavelet Theory and its Application*. Kluwer Academic Publishers, Bonston, 1993.

[4] D. Keim and M. Heczko. Wavelets and their applications in databases. Tutorial Notes of ICDE 2001, 2001.

[5] D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. The MIT Press, 2001.

[6] J. Han and M. Kamber. *Data Mining: Concepts and Techniques* Morgan Kaufmann Publishers, 2000.

[7] I. Bruha and A. F. Famili. Postprocessing in machine learning and data mining. *SIGKDD Exlporations*, 2000.

[8] C. K. Chui and J. Lian. A study of orthonormal multiwavelets. *Applied Numerical Mathematics: Transactions of IMACS*, 20(3):273–298, 1996.

[9] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989

[10] R. Polikar. The wavelet tutorial. Internet Resources: http://engineering.rowan.edu/ polikar/ WAVELETS/WTtutorial.html.

[11] I. Daubechies. *Ten Lectures on Wavelets*. Capital City Press, Montpelier, Vermont, 1992

[12] D. Gunopulos. Tutorial slides: Dimensionality reduction techniques. In DIMACS Summer School on Data Mining,August 2001.

[13] F. Abramovich, T. Bailey, and T. Sapatinas.Wavelet analysis and its statistical applications. *JRSSD*, (48):1–30, 2000.

[14] D. H. Fisher. Iterative optimization and simplification of hierarchical clusterings. Technical Report CS-95-01, Vanderbilt U., Dept. of Comp. Sci., 1995.

[15] P. Smyth. Probabilistic model-based clustering of multivariate and sequential data. In *Proceedings of Artificial Intelligence and Statistics*, 1999.

[16] S. Zhu, T. Li, and M. Ogihara. CoFD: An algorithm for non-distance based clustering in high dimensional spaces. In *Proceedings of 4th International Conference DaWak 2002*, number 2454 in LNCS, pages 52–62, Aix-en-Provence, France, 2002. Springer.

[17] C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, and J. S.Park. Fast algorithms for projected clustering. pages 61–72, 1999.

[18] S. Guha, R. Rastogi, and K. Shim. CURE: an efficient clustering algorithm for large databases. In *Proceedings of ACM SIGMOD*, pages 73–84, 1998.

[19] S. Guha, R. Rastogi, and K. Shim. ROCK: A robust clustering algorithm for categorical attributes. *Information Systems*, 25(5):345–366, 2000.

[20] V. Ganti, J. Gehrke, and R. Ramakrishnan. CACTUS – clustering categorical data using summaries. In *Knowledge Discovery and Data Mining*, pages 73–83, 1999.

[21] M. Berger and I. Rigoutsos. An algorithm for point clustering and grid generation. *IEEE Trans. on Systems, Man and Cybernetics*, 21(5):1278–1286, 1991.

[22] G. Sheikholeslami, S. Chatterjee, and A. Zhang. WaveCluster: A multi-resolution clustering approach for very large spatial databases. In *Proc. 24th Int. Conf. Very Large Databases, VLDB*, pages 428–439, 1998.

[23] V. Castelli and I. Kontoyiannis.Wavelet-based classification: Theoretical analysis. Technical Report RC-20475, IBM Watson Research Center, 1996.

[24] V. Castelli and I. Kontoyiannis. An efficient recursive partitioning algorithm for classification, using wavelets. Technical Report RC-21039, IBM Watson Research Center, 1997.

[25] G. Sheikholeslami, A. Zhang, and L. Bian. A multiresolution content-based retrieval approach for geographic images. *GeoInformatica*, 3(2):109–139, 1999.

[26] M.B. Sanjay Pande and P.Nagabhushan. An algorithmic model for exploratory analysis of trace elements in cognition and recognition of neurological disorders. Ph.D. Thesis. Department of studies in computer Science, University of Mysore. August,2004.

[27]Diday.E. Symbolic Data Analysis : a theory and tool for Data Mining, IFCS, Rome,1998

[28]Diday.E.Symbolic Data Analysis:some in and out,KESDA,1998

R. Pradeep Kumar (1977) completed his bachelors in Electrical Engineering from Amrita Institte of Technology in 1999, India and Masters in Computer Sceince & Engineering from SJCE Mysore , India during 2001. His Ph.D. is from department of studies in computer science, University of Mysore, India.
He has been an active academician and researcher for the past 6 years. His areas of interest include Data and Knowledge Engineering, Image and Video Processing and Computational Intelligence. He is professional member of ACM.

Nagabhushan. P completed his bachelors in Electrical Engineering from SJCE Mysore, India and Masters from BITS Pilani, India. He obtained his Ph.D. from University of Mysore, India.
He has over 300 publications in reputed journals and conferences. He is an active researcher in the areas pertaining to Pattern Recognition, Image and Video Processing, Symbolic Data Analysis and Computational Intelligence. He is a Fellow of Institution of Engineers.