Wavelets or wavelet analysis or the wavelet transform refers to the representation of a signal in terms of a finite length or fast decaying oscillating waveform known as the mother wavelet. This waveform is scaled and translated to match the input signal. The wavelet transform coefficients has been used as an index of similarity between a function $f(t)$ and the corresponding wavelet, in the fields of pattern recognition and knowledge discovery. In these fields, the coefficients are generated to acquire a set of features. This chapter attempts to make two important contributions in using wavelets for temporal data analysis. First, this chapter proposes a novel transform for generating wavelet like coefficients by using a conventional similarity measure between the function $f(t)$ and the wavelet i.e. the wavelet coefficients are indices of similarity, and the proposed transform is an alternate method to generate a normalized set of similarity indices, whose characteristics are similar to that of wavelet coefficients. WaveSim Transform is demonstrated with an application in temporal data clustering. The concept is extended to trend clustering application in time series. Second, this chapter presents a technique for visualizing an entire time series as a point at multiple resolutions. A methodology for representing a time series (or temporal signal) as histograms at different resolutions using wavelet transform is proposed. Then a regression line is fitted on to the cumulative histogram and the line is then expressed as a point in the Hough space. The novel contributions are demonstrated through SOX dataset, which simulates the scenario corresponding to temporal signals.
4.1 Introduction

Wavelet analysis is becoming a common tool for analyzing localized variations of power within temporal signal. By decomposing a temporal signal into time-frequency space, one is able to determine both the dominant modes of variability and how those modes vary in time. Wavelet is a powerful tool, which we can use not only to study properties of signals or functions, but also to represent them. Being able to represent signals concisely allows us to determine which parts of the signal are truly important, and which parts are details we can live without. The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. Wavelet is an excellent tool in the field of data compression. Other applied fields that make use of wavelets include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, and pure mathematics applications such as solving partial differential equations.

But the mathematical complexity depicted in the computation of wavelet transform by many literatures has always inhibited researchers in many areas from understanding and utilizing it for data analysis in different application domains. In pattern recognition and knowledge discovery applications, the wavelet transform coefficients were used as a similarity index and the descriptors extracted from these coefficients were providing efficient representation of samples. In this chapter with a mindset focused on PR research, a reverse approach for computing the wavelet domain coefficients in the name of WaveSim Transform is proposed. The coefficients generated as a result of this transform are a normalized set of coefficients whose characteristics are same as that of wavelet transform. These normalized set of coefficients can be used for multiresolution data analysis in the WaveSim domain.
The important emphasis throughout this chapter is to seed a possibility of using the conventional similarity measures between a set of features extracted from the wavelet and the function \( f(t) \) spanning the wavelet range at an instant 'b', for generating the coefficients. These coefficients are termed as WaveSim Transform coefficients as these are the coefficients generated out of a similarity measure between the wavelet and the range of \( f(t) \) spanning over the wavelet at a particular scale. The results obtained are interesting. This idea is oriented in section 4.3 after a brief interesting interpretation of CWT in section 4.2.

The multi resolution coefficients obtained from wavelet transform or wavesim transform could be further modeled through histogram and regression lines for visualization. This idea is conceived in section 4.4.

4.2 Interpretation of CWT

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

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

(4.1)

where \( a \) and \( b \) are real and * denotes complex conjugation (Grossman, Morlet, 1984; Daubechies, 1990; Daubechies, 1992). Thus, the wavelet transform is a function of two variables. Both \( f(t) \) and \( \psi(t) \) belong to \( L^2(\mathbb{R}) \), 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:

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

(4.2)

CWT has a cross correlation interpretation as well

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

(4.3)
In other words, the CWT is the cross-correlation at lag $b$ between $f(t)$ and the wavelet dilated to scale factor $a$. Using the Cauchy-Schwartz inequality

$$|W(a,b)|^2 \leq ||f(t)||^2 ||\psi_{a,b}(t)||^2$$

(4.4)

This in turn implies that $W(a,b)$ always exists because the function and the wavelet have finite norms. Equality holds in the relation if and only if

$$\psi_{a,b}(t) = \propto f(t)$$

(4.5)

for some scalar $\propto$. Therefore if there actually is a pair of values $(a,b)$ for which the equation is satisfied, then the scalogram $|W(a,b)|^2$ achieves its maximum at this pair of values.

A measure of the separation between $f(t)$ and $\psi_{a,b}(t)$ is provided by $||f(t) - \psi_{a,b}(t)||^2$. We have

$$||f(t) - \psi_{a,b}(t)||^2 = ||f(t)||^2 + ||\psi_{a,b}(t)||^2 - 2Re[W(a,b)]$$

(4.6)

where Re[ ] stands for “real part of”. $\psi_{a,b}(t)$ is closest to $f(t)$ for that value of the pair $(a,b)$ for which the left hand side is minimized. Of the quantities on the right hand side, only Re[$W(a,b)$] varies with $a$ and $b$. It has a negative sign associated with it. Thus the maximum of Re[$W(a,b)$] corresponds to the minimum value of the left-hand side, yields the values of $a$ and $b$ that provide the best positive correlation between $f(t)$ and $\psi_{a,b}(t)$.

Similarly, working with $||f(t) + \psi_{a,b}(t)||^2$, which provides a measure of closeness between $f(t)$ and $-\psi(t)$, we see that the minimum value of Re[$W(a,b)$] yields $a$ and $b$ for which there is the best negative correlation between $f(t)$ and $\psi_{a,b}(t)$. Thus the maximum value of Re[$W^2(a,b)$] points to the best possible match between a dilated and translated version of the mother wavelet and the input signal $f(t)$. Local maxima usually disclose local features in the signal that correlate well with the wavelet for values of dilation, and shift corresponds to the location of such maxima (Raghuveer, Ajith, 2000).

It is these interpretations which make Wavelet Transform an astonishing transform having larger orientation towards pattern recognition and knowledge mining because details/features of $f(t)$ can be captured at multiple scales/resolutions in the transformed domain. We shall now start considering the wavelet function $\psi_{a,b}(t)$ as a template. Now we interpret wavelet transform as a process of moving the template...
over the function $f(t)$ and measuring the similarity between the template and a segment of $f(t)$ of width ‘$a$’ at a time instant ‘$b$’ . Thus $W^2(a,b)$ is the value of similarity measure between the template $\psi_{a,0}(t)$ and a segment of $f(t)$. With this as the interpretation we like to provide a new direction to the interpretation of wavelet transform by defining a new transform called as WaveSim transform. In WaveSim transform we suggest a similarity measure between the template and the segment of $f(t)$ at time instant ‘$b$’ and it is observed that the coefficients generated with such a kind of similarity measure are normalized and thus ready to use as inputs to most classifiers.

### 4.3 Wavesim Transform

The WaveSim(WS) transform of $f(t)$ with respect to a wavelet $\psi_{a,b}(t)$ is defined as

$$WS(a,b) = Sim(f(t), \psi_{a,b}(t))$$

(4.7)

where $Sim(.,.)$ is the similarity measure computed as illustrated in figure 4.1 with a Haar Wavelet. The amplitude of $\psi$ is $\max(f(t))$.

![Fig. 4.1a](image-url)
The WaveSim coefficient is computed as follows

\[
\text{Sim}(f(t), \psi(t)) = \frac{1}{2}((A_3 - A_5)/A_1 + (A_4 - A_6)/A_2)) \tag{4.8}
\]

where \((A_3/A_1)\) is the similarity measure between the positive half of the wavelet and \(f(t)\) in the positive scale range, \((A_5/A_1)\) is the dissimilarity measure between the positive half of the wavelet and \(f(t)\) in the positive scale range.

Likewise \((A_4/A_2)\) is the similarity measure between the negative half of the wavelet and \(f(t)\) in the negative scale range, \((A_6/A_2)\) is the dissimilarity measure between the negative half of the wavelet and \(f(t)\) in the negative scale range. \(\frac{1}{2}\) helps in normalizing the coefficient values to lie in the interval \([-1,1]\).
4.3.1 WaveSim Transform: Analysis

WaveSim Transform results in the generation of a normalized set of coefficients with values from -1 to 1. The coefficients are generated at different scales as computed in the case of continuous wavelet transform with a varying translation parameter ‘b’. It has always been convenient to handle normalized values in pattern recognition and various soft computing fields. On plotting the coefficients obtained from WaveSim Transform and comparing it with that of continuous wavelet transform coefficients, it is interesting to note that, the behavior of both the plots are alike. The correlation between these two sets of coefficients is 1.

This leads to many interesting inferences. The coefficients generated by WaveSim Transform is a similarity coefficient between the template/wavelet and the function f(t). It is based on the area occupied/spanned by the function f(t) over a template/wavelet. It is not just another kind in the interpretation of Wavelet Transform. This Transform could open up an arena where a set of feature points in f(t) spanning over a wavelet/template can be compared with a set of feature points extracted from the wavelet. The distance between these two sets of feature points is expected to generate coefficients whose correlation with respect to wavelet coefficients would be almost 1. The area based similarity measure proposed in this approach can be replaced by any other similarity measure.

The results obtained for different signal inputs are provided for illustrating that, both Wavelet Transform coefficients and WaveSim transform coefficients exhibit similar behavior/ characteristics. The results have been computed at scale = 10. The results were verified for a range of scales. The different cases for which it was tested are

Case 1: Input signal is a set of 200 samples with equal values (f(t) is a horizontal straight line)
Case 2: Input signal is a set of 200 samples with the first 100 samples holding a value 20 and the second 100 samples holding a value 80.

Case 3: Input signal is a set of 200 samples linearly increasing from 1 to 200. (f(t) is a straight line)
**Case 4:** Input signal is a set of 200 samples with a curved structure.

For all the above cases the correlation coefficient between the WaveSim Transform coefficients and the Wavelet Transform coefficients is 1. WaveSim transform is further investigated in chapter 5 for exploratory analysis of temporal sequences.
following sections crisply demonstrate the strength of wavesim and wavelet transforms in temporal data analysis.

4.3.2 WaveSim Transform for Temporal Data Clustering

Given a source of temporal / time series data, such as the stock market or the monitors in an intensive care unit, there is high utility in determining whether there are qualitatively different regimes in the data and then in characterizing those regimes. For example, one might like to know whether the various indicators of a patient's health measured over time are being produced by a patient who is likely to live or one that is likely to die. In this case, there is a priori knowledge of the number of regimes that exist in the data (two), and the regime to which any given time series belongs can be determined post hoc (by simply noting whether the patient lived or died) (Time Oates, Laura, 1999). However, these two pieces of information are not always present.

Let $S$ denote a time series spanning $n$ time steps such that $S = \{ S_t | 1 \leq t \leq n \}$ where $S_t$ is the value at time instant $t$. Given a set of $m$ time series, we want to obtain, in an unsupervised manner, a partition of these time series into subsets such that each subset corresponds to a qualitatively different regime. In this section, we propose a method for extracting WaveSim features from a time series data for clustering. The clustering can be performed at multiple resolutions with WaveSim features extracted at multiple scales. As in case of Wavelet Transform the coefficients at the dyadic scales (scales at 2, 4, 8, 16, 32, 64 ...) are unique enough to extract the features and these features give us the capacity to view the clusters at multiple resolutions.

Algorithm for WaveSim feature extraction

**Step 1:** Consider a temporal data set $D$ with $m$ multivariate time series. A sample time series is shown in figure 4.3.

**Step 2:** Quantize the time line into $k$ intervals with each interval spanning a time interval of $I$ time units as illustrated in figures 4.3 and 4.4.
Fig 4.3 A sample time series and the intervals

Time span of each temporal signal

\[ \begin{align*}
I_1 & \quad I_2 \quad I_3 \quad I_4 \quad I_5 \quad I_6 \quad I_7 \quad I_8 \quad I_9 \quad I_{10} \\
\text{No of intervals} &= 10 \quad (\text{Sample Case})
\end{align*} \]

Fig 4.4 Quantized time span with 10 intervals

Step 3: For each signal data \( S \) apply WaveSim Transform with a Haar Wavelet at Scale \( S_{CX} \) where \( X \) is a dyadic scale value.

Corresponding WaveSim Coefficients at scale \( S_x \)

\[ \begin{align*}
I_1 & \quad I_2 \quad I_3 \quad I_4 \quad I_5 \quad I_6 \quad I_7 \quad I_8 \quad I_9 \quad I_{10} \\
\end{align*} \]

Fig 4.5 WaveSim coefficients over the time line

Step 4: Sort the WaveSim Coefficients in descending order and capture the corresponding indices of the sorted array.

Sorted WaveSim Coefficients

\[ \begin{align*}
I_1 & \quad I_2 \quad I_3 \quad I_4 \quad I_5 \quad I_6 \quad I_7 \quad I_8 \quad I_9 \quad I_{10} \\
\end{align*} \]

Fig 4.6 Sorted WaveSim coefficients over the time line

Step 5: Collect the first \( P \) number of indices into an array \( A \), divide the entire array with value \( I \) and ceil it. When \( P \) was set to value between 40 and 100, it resulted in
considerably better performance during experimentation. Now generate a Histogram which gives the frequency count of the indices values falling into the $k$ intervals. So each temporal Signal can be represented by $k$ features.

Step 6: For the obtained feature set a clustering technique is applied and the different regimes are viewed.

Step 7: Step 3 to Step 6 are repeated at different dyadic scales starting from $S_{cX} = 2$ until $S_{cX} < n/2$ where $n$ is the length of the temporal signal.

4.3.3 Experimentation Results
The proposed transform and the feature set extracted from it were first tested on the temporal dataset shown as plots in figure 4.8. As can be observed from the figure, there exist 3 global classes. If the temporal samples are observed from different granularities or scales the association between the samples varies interestingly. The dendrograms shown in figure 4.9 depict the interesting associations/clusters at different scales. This leads to multi resolution knowledge discovery where novel patterns get uncovered at different granularities. The feature sets were extracted by setting $k = 10 \Rightarrow I = 200/10 = 20$ time units and $P = 50$. 

Fig 4.7 Mapping the Indices of sorted WaveSim coefficients to the Corresponding interval they belong
WAVESIM TRANSFORM: A NOVEL PERSPECTIVE OF WAVELET TRANSFORM FOR TEMPORAL KNOWLEDGE MINING

Fig. 4.8 Input Samples

Fig 4.9a Dendrogram at Scale 2

Fig 4.9b Dendrogram at Scale 4

Fig 4.9c Dendrogram at Scale 8

Fig 4.9d Dendrogram at Scale 16
Discussions:
There are 3 major classes in the samples that are considered (figure 4.3). The first 8 samples belong to ‘8’s group, 2\textsuperscript{nd} 8 samples belongs to ‘O’s group and 3\textsuperscript{rd} 8 samples belong to ‘X’s group.

At Scale 2 : \{4, 6\}, \{13\}, \{8,21\}, \{20\}, \{22\} samples pop out because of its unusual shape when observed from a finer perspective.

At Scale 4 : \{13\} pops out (Abnormal Shape)

At Scale 8 : \{20\} pops out (Abnormal Shape)

At Scale 16: \{12\} \{20\} \{22,23\} pops out (Abnormal Shape)

At Scale 32: \{20\} pops into the ‘8’s category when the observation becomes more and more approximate. Other samples are intact.

At Scale 64: All samples intact except \{20\}.

So analyzing the samples at different resolutions pops out the noisy samples. But it is very important to note that they are not really noisy samples because in reality they belong to one of the category. It can be clearly observed that, at some resolutions these samples take a place in their own group at least once, as in case of 20 at scale 4.

So analysis of samples at multiple resolution can always result in interesting knowledge discovery. Chapter 6 and 7 deal with multiresolution knowledge discovery in more depth.
4.4 WaveSim Histograms for Time Series Trend Clustering

Time series is a record of the values of any fluctuating quantity measured at different points of time. One characteristic feature which distinguishes time series data from other types of data is that, in general, the values of the series at different time instants will be correlated. Application of time series analysis techniques in temporal data mining is often called Time Series Data Mining. A great deal of work has been done in recent past into identifying, gathering, cleaning, and labeling the data, into specifying the questions to be asked of it, and into finding the right way to view it to discover useful temporal patterns. In this section we propose a model for time series whole clustering based on wavesim histograms. The notion of whole clustering here is similar to that of conventional clustering of discrete objects. Given a set of m multivariate time series, we want to obtain, in an unsupervised manner, a partition of these time series into subsets such that each subset corresponds to a particular trend. In this section we propose a method for extracting positional histogram features, WaveSim summary histogram features and WaveSim energy histogram features from a time series data for whole clustering. These histogram features are used for hierarchical clustering of time series segments. Usually a time series would belong to any one of the following trend categories:

1. Normal
2. Cyclic
3. Increasing trend
4. Decreasing trend
5. Upward shift
6. Downward shift

A sample schematic of the above categories is shown in figure 4.10 (courtesy Eamonn Keogh). Further the time series will have their own similarities based on their behavior. So we first achieve trend clusters through positional histogram features and WaveSim summary histograms, and then we achieve behavioral similarity clusters based on WaveSim energy histogram features.
Fig 4.10 (A) Downward Trend. (B) Cyclic. (C) Normal. (D) Upward Shift. (E) Upward Trend. (F) Downward Shift

4.4.1 Extracting Positional Histogram Features from Time Series

The idea behind a positional histogram feature is very simple. Divide the time series \( S \) into 3 equal segments \( S_1, S_2 \) and \( S_3 \). For each of these segments generate a histogram for the normalized set of time series magnitudes. These positional histograms are capable of achieving first level clusters by resulting in 3 regimes. The first regime is normal & cyclic group, the second being downward trend and downward shift group and the third one being the upward shift and upward trend group. This is possible because

i. The histogram for all three segments \( S_1, S_2 \) and \( S_3 \) will have a uniform spread in case of cyclic and normal. It is obvious because the values all along the segment vary between 0 and 1 because the time series is normalized.

ii. In case of upward trend and upward shift \( S_1 \) will result in a left skewed histogram, \( S_2 \) will exhibit a centrally spread histogram and \( S_3 \) will be a right skewed histogram. It is because the left segment \( S_1 \) will have values occupying the lower ranges of \([0 \ 1]\), \( S_2 \) will occupy the middle ranges of \([0 \ 1]\) and \( S_3 \) will occupy the higher ranges of \([0 \ 1]\).
iii. In case of downward trend and downward shift, S1 will result in a right skewed histogram, S2 will be a centrally spread histogram and S3 will result in a left skewed histogram. It is because the left segment S1 will have values occupying the higher ranges of [0 1], S2 will occupy the middle ranges of [0 1] and S3 will occupy the lower ranges of [0 1].

![Fig 4.11 Plots of time series showing different trends](image)

![Fig 4.12 Positional histograms of the time series shown in fig 4.11](image)

### 4.4.2 Extracting WaveSim Summary Histogram Features from Time Series

*Algorithm for WaveSim summary feature extraction:*

**Step 1:** Consider a time series data set D with m multivariate time series.

**Step 2:** For each time series signal S apply WaveSim Transform with a Haar Wavelet (Any wavelet can be used) at Scale ScX where X is a dyadic scale value.

**Step 3:** Now generate a Histogram of the WaveSim coefficients at scale Sx. We are already aware that the coefficients of WaveSim transform fall into a range [-1 +1].
Step 4: Repeat Step 2 to Step 3 at different dyadic scales starting from $S_{cX} = 2$ until $S_{cX} < n/2$ where $n$ is the length of the temporal signal.

This is an attempt to analyze the time series objects with respect to the global information and local information conveyed by the time series. Here WaveSim histograms are generated at multiple dyadic scales. We exploit the global information or low resolution factors in the time series with wavelets of higher scales whereas the local information or finer resolution factors are extracted with a smaller scale wavelet. Thus the summary histogram features extracted are a summary of local and global information about the time series extracted at multiple resolutions. Thus we obtain $p = \log_2 S_{cX}$ number of WaveSim summary histograms corresponding to ‘$p$’ dyadic scales. So each time series is associated with ‘$p$’ number of histograms. This helps us in categorizing the upward group into two clusters i.e. upward trend and upward shift, the downward group into downward trend and downward shift and also to separate out the normal and cyclic groups.

4.4.3 Extracting WaveSim Energy Histogram Features

Once the time series segments have been clustered based on their trend characteristics now we try to extract the intra cluster association between samples through WaveSim energy histograms. This histogram is based on the energy concentration of WaveSim coefficients in those segments of the time series whenever the similarity between the wavelet and the $f(t)$ segment spanned by the wavelet is high. This histogram classifies the samples based on the time instant of upward/downward shift, rate of upward trend or downward trend, the rate of cycles in a cyclic trend and the behavior of normal trends etc. The algorithm for extracting WaveSim energy histogram is exemplified below:

Step 1: Consider a time series data set $D$ with $m$ time series.

Step 2: Quantize the time line with ‘$n$’ time instants into $k$ intervals with each interval spanning a time interval of $I$ time units.

Step 3: For each temporal signal $S$ apply WaveSim Transform with a Haar Wavelet (Any wavelet can be used) at Scale $S_{cX}$ where $X$ is a dyadic scale value.

Step 4: Sort the WaveSim Coefficients in descending order and capture the corresponding indices of the sorted array.
Step 5: Collect the first $P$ (where $40 < P < 100$) number of indices and determine the interval into which each index falls. Now generate a Histogram which gives the energy sum of the coefficients corresponding to the indices values falling into the $k$ intervals. So each temporal Signal is now represented by a distribution or a histogram with $k$ number of bins.

Step 6: Repeat Step 3 to Step 5 at different dyadic scales starting from $Sc_x = 4$ until $Sc_x < n/2$ where $n$ is the length of the temporal signal.

Thus even in this case we obtain $p = \log_2 Sc_x$ number of WaveSim energy histograms corresponding to ‘$p$’ dyadic scales. So each time series is associated with ‘$p$’ number of histograms.

Hence in case of positional histogram we get three histogram features corresponding to each segment. WaveSim summary histograms and WaveSim energy histograms are generated at ‘$p$’ resolutions which imply that we obtain ‘$p$’ number of histogram features for each time series object. We use the AB regression based histogram distance introduced in section 2.3.6 for generating distance matrix from the histogram features.

4.4.4 Histogram based Hierarchical Clustering Technique

With the histogram features thus obtained we define a 4 level hierarchical tree structure for visualizing the cluster procedure as shown in figure 4.13. In level 1 the root node consists of the entire time series data set. From these data set a set of positional histogram features are extracted and a distance matrix is computed by using the AB regression based distance measure. With this distance matrix we get 3 dominant clusters $C_1$ (normal, cyclic), $C_2$ (Upward trend, Upward Shift) and $C_3$ (Downward trend, Downward shift). Then we extract a set of WaveSim summary histogram features as described earlier and we compute the distance matrix with which we could separate the two dominant clusters in each of the level 2 clusters thus obtaining $C_{11}$, $C_{12}$, $C_{21}$, $C_{22}$, $C_{31}$ and $C_{32}$ as labeled in figure 4.13. Then we use the WaveSim energy histogram features for categorizing the level 3 clusters into further local associations.
4.4.5 Experimentation Results and Analysis
We have considered the synthetic control chart data set generated by Eamonn Keogh consisting of 600 time series samples. This is a good data set to test time series clustering (and classification) algorithms because Euclidean distance will not be able to achieve perfect accuracy. In particular the following pairs of classes will often be confused (Normal/Cyclic) (Decreasing trend/Downward shift) and (Increasing trend/Upward shift). Keogh demonstrates that the Derivative Dynamic Time Warping is capable of categorizing these time series sequence. And this is the next model which categorizes these time series perfectly. The technique proposed by us based on positional and WaveSim histograms would always result in stable clusters while maintaining the global and local association in the clusters. Out of the 600 samples 1-100 are normal, 101-200 are cyclic, 201-300 are increasing trend, 301 to 400 are
upward shift, 401-500 are decreasing trend and 501-600 are downward shift. The results obtained with the positional histogram are shown in figure 4.14. It portrays the 3 dominant level-2 clusters. Then the clusters C1, C2 and C3 are further classified using the WaveSim summary histogram features to obtain the clusters C11, C12, C21, C22, C31 and C32. Figures 4.15, 4.16 and 4.17 show the cluster results.

Fig 4.14 Positional histogram feature based clusters

Fig 4.15 WaveSim summary histogram feature based clusters: Cyclic and Normal
WAVESIM TRANSFORM: A NOVEL PERSPECTIVE OF WAVELET TRANSFORM FOR TEMPORAL KNOWLEDGE MINING

Clusters obtained using WAVESIM Summary Histogram Features

Upward Trend

C22

Upward Shift

C21

Fig 4.16 WAVESIM summary histogram feature based clusters: U.Trend and U.Shift

Clusters obtained using WAVESIM Summary Histogram Features

Downward Trend

C32

Downward Shift

C31

Fig 4.17 WAVESIM summary histogram feature based clusters: D.Trend and D.Shift
The clusters obtained at level 3 are highly stable and the important aspect is that the local associations within the clusters are still maintained. Further clustering of $C_{11}$, $C_{12}$, $C_{21}$, $C_{22}$, $C_{31}$ and $C_{32}$ with positional energy histograms signify the local associations and the associations depend on the rate of increasing trend or decreasing trend, the time instant of the upward or downward shift, the rate of cycles etc.

4.5 A Novel Model for Visualising Time Series as Points at Multiple Resolutions

Wavelet transforms is also a powerful concept for characterizing and visualizing information contained in data at multiple resolutions. In this section we propose a methodology for characterizing short time series data at multiple resolutions by applying waveSim transform. The coefficients are modeled as histograms. These histograms are transformed into regression lines which are further represented in a point space. Thus the entire time series is represented as a single point and paves way for visualizing time series clusters in a scatter space.

4.5.1 WaveSim Feature Extraction

The following procedure is used for extracting the wavelet histogram features from a set of time series:

1. **Step 1**: Consider a temporal data set $D$ with $m$ multivariate time series.
2. **Step 2**: Quantize the time line into $k$ intervals with each interval spanning a time interval of $I$ time units.
3. **Step 3**: For each temporal data $S$ apply WaveSim transform with a Haar Wavelet at Scale $S_{cX}$ where $X$ is a dyadic scale value.
4. **Step 4**: Sort the waveSim coefficients in descending order and capture the corresponding indices of the sorted array.
5. **Step 5**: Collect the first $P$ (where $40<P<100$) number of indices into an array $A$, divide the entire array with value $I$ and ceil it. Now generate a Histogram which gives the frequency count of the indices values falling into the $k$ intervals. So each time series signal can be represented by a distribution or a histogram with $k$ number of bins.
6. **Step 6**: Repeat Step 3 to Step 5 at different dyadic scales starting from $S_{cX} = 2$ until $S_{cX} < n/2$ where $n$ is the length of the temporal signal.
The resultant is a dataset with each time series represented by a wavesim histogram corresponding to scale X. Thus for ‘p’ number of scales we obtain ‘p’ such dataset.

4.5.2 Histogram as a Point in Hough Space

The histograms representing the time series are transformed into regression lines as illustrated in section 2.3.1 in chapter 2. Then these regression lines are transformed to a point in the Hough space through a Hough transform approach.

Let the equation of the regression line denoted by A be given by

\[ y_1 = mx_1 + c_1 \]  (4.9)

Therefore the equation of a line B drawn perpendicular to the regression line and passing through the origin is given by

\[ y_2 = (-1/m) x_2 \]  (4.10)

For line B we compute the amount of inclination with respect to the x axis and it is denoted by \( \theta \). Also we compute \( \rho \) which is the displacement of line A from the origin. Now we plot the point \( (\theta, \rho) \) in the Hough space where \( \theta \) is the x axis and takes value from 0 to 180 degrees and \( \rho \) form the y axis. The \( \theta \) and \( \rho \) for the line pairs A and B are computed as follows

\[ \rho = c_1 / (\sqrt{1+m^2}) \]  (4.11)

\[ \theta = \tan^{-1}(-1/m) \text{ if } (-1/m) \geq 0 \]
\[ = 180 - \tan^{-1}(-1/m) \text{ if } (-1/m) < 0 \]  (4.12)
Thus the histogram is represented as a point in the Hough space which implies that an entire time series is characterized by a single point.

4.5.3 Experimentation Results and Discussion

The proposed technique and the feature set extracted from it were first tested on the temporal dataset shown as a plot in figure 4.19. It can be observed from the figure 4.19, that there exist 3 global classes. If the temporal samples are observed at different granularities or scales the association between the samples varies interestingly. The dendrograms depict the interesting associations/clusters at different scales. This leads to multi resolution knowledge discovery where novel patterns get uncovered at different granularities. The feature sets were extracted by setting $k = 10 \Rightarrow I = 200/10 = 20$ time units and $P = 50$. The clustering is performed based on AB Regression based distance between the histograms. Once the histogram is transformed into a regression line it can be visualized as a point in the slope-intercept point space or Theta-Rho Hough space. It can be observed that the scatter plots characterize the dendrograms very closely.

![Fig 4.19 Input Samples](image-url)
Fig 4.20a Histogram features obtained at scale 64

Fig 4.20b Normalized Cumulative Histogram Features

Fig 4.20c Regression lines fitted to the normalized histograms
Fig 4.20d Dendrogram obtained through complete linkage clustering using AB distance measure

Fig 4.20e Scatter plot – Slope Vs Intercept
Discussion:

There are 3 major classes in the samples that are considered (figure 4.19). The first 8 samples belong to '8's group, 2\textsuperscript{nd} 8 samples belong to 'O's group and 3\textsuperscript{rd} 8 samples belong to 'X's group. In the dendrogram shown in figure 4.20d we can observe sample 12 popping out completely which is also same in the scatter plots in figure 4.20e and 4.20f. Sample 5 is mixed with 'X' group and sample 20 mixing with the first group is clearly depicted by the scatter plots. Sample 13 mixing with the first group is also prominent in the scatter diagram which is also visible in the dendrogram. Thus if a time series can be characterized as a distribution type symbolic histogram object it is always possible to visualise it as a point in the point space.
Fig 4.21a Histogram features obtained at scale 32  Fig 4.21b Normalized Cumulative Histogram Features

Fig. 4.21c: Regression lines fitted to the normalized histograms
Fig. 4.21d: Dendrogram obtained through complete linkage clustering using AB distance measure

Fig. 4.21e: Scatter plot - Slope Vs Intercept
4.6 Conclusion
This chapter attempted to propose a novel transform which can be used for multiresolution analysis. Novel similarity measures that would be capable of generating coefficients whose characteristics may be similar or better than that of Wavelet Transform coefficients could be a prospective future direction. Chapter 5 deals with one such attempt. The attempt would enhance curiosities to simplify approaches for multiresolution analysis without much mathematical complexities. Although lot of research has gone into time series data clustering, there is no much attempt made in terms of visualizing the time series clusters in the point space. Most of the results produced by time series researchers are provided with the help of dendrograms. Thus this attempt would seed a possibility of visualizing time series in a point space. The basic methodology should be to efficiently summarize and represent a time series with appropriate symbolic objects and then transform the symbolic objects into a point space. Hence there is a lot of scope for further research especially in case of visualizing multichannel time series data in the point space. The proposed attempt will be a breakthrough for time series cluster visualization.
It is with such novel explorations many new-fangled models are proposed for temporal signal analysis in the following chapters.