Turning Function Based Distance Measure for Searching Similar ECG trajectories

Gajanan S. Gawde
Goa Engineering College
Farmagudi-Ponda, Goa-India

Jyoti D. Pawar
Goa University
Taleigao-Panaji, Goa-India

ABSTRACT: *Medical field is becoming more and more challenging due to large number of heart problems faced by the people. Almost all doctors are always trying their best to give a best treatment to a patient so that the patients get recovered from the heart problem. The heart patient is carefully diagnosed by analyzing ECG signals. Many researchers have proposed dierent distance measures to compare ECG trajectories for similarity. Dynamic Time Warping(DTW), Edit Distance with Real Penalty(ERP), Edit Distance with Real Sequence (EDR), Longest Common Subsequence(LCSS) are used to compare time series trajectories for similarity. The existing distance measures are computing similarity distance based on a proximity distance. There are three main drawbacks faced by the existing distance measures. The first drawback is, ECG trajectories are not compared based on a shape feature. The second problem is, existing edit distance measures are not rotation invariant. The third problem is, they are very sensitive towards noise. Due to these problems, the performance of the existing distance measures is comparatively low.*

In this paper, we have proposed Turning Function Based (TFB) distance measure to compares ECG trajectories based on a shape feature. TFB distance measure is based on a angular distance of the ECG trajectories. The angular distance captures the shape feature of the ECG trajectory. We have carried out experimental study on the real time and synthetic datasets. Experimental results reveal that our proposed TFB distance measure compares the trajectories based on the shape feature. Further, our experimental results reveal that, TFB distance measure supports rotation invariant property and very robust to the noise.

Key words: Turning Function Based Distance Measure, ECG Trajectories, Shape Based Matching, Rotation Invariant Distance Measure

Received: 18 September 2016, Revised 23 October 2016, Accepted 3 November 2016

1. Introduction

Now a days, a large number of people are suering from heart problem. There is a need to provides a good treatment to the heart patients. The heart functioning of the patient is analyzed through ECG signals. In normal method, the doctors are taking a printout of ECG using machine and the ECG graph obtained is compared for abnormal behavior or some kind of patterns which would suggest that there is some problem with a heart functioning. The checking of the ECG graph is done manually and there is always chance of skipping some vital information in this process by the doctors. This manual checking of ECG is sometimes very dangerous, since there is every possibility that doctors might skip or overlooked some interesting information which is linked to the heart patient. Most of time doctors are busy with their work and they are hardly getting any time to take rest. Due to extra work load and continuous work, they may not able to analyze the ECG signal systematically.

To provides good treatment to the patient, good diagnosis is very important. The main cause of the problem need to be identified and then only doctor can prescribe good medicine. Manual way of checking ECG graph is not proper and the ECG signals need to be analyzed deeply. Then only doctors might get good clue about the illness and accordingly doctor can give good treatment in the form of medicines and other pharmaceutical drugs. Heart disease is one of the most dangerous disease many patients are suering. The patients are always worried about the treatment given by the doctors. Sometimes it might happen due to wrong diagnosis patient are given wrong medicines which might create side eect in the body of the patient. Hence there is need to take utter care while doing diagnosis and identifying the right cause of the illness. This would give lot of confident in the minds of the patients and also doctors would get satisfaction from their work.

The time series trajectories of ECG of the patients are compared using edit distance measures. These edit distance measures are working purely on a proximity distance. The proximity distance computed using edit distance measures is not able to identify similar trajectories accurately. The ECG trajectory has some kind of shape and this shape feature can be used to compares trajectories for similarity. In this paper, we have proposed Turning Function Based distance measure which compares ECG trajectories based on the shape feature.

2. Background

Time series trajectories are gaining lot of importance due to wide range of applications. The ECG of a patient is a time series trajectory which contain important information about heart functioning. ECG time series trajectories are analyzed deeply to identify any abnormal behaviour of heart functioning of a patient. The parameters of ECG signals are changing with respect to the time axis.

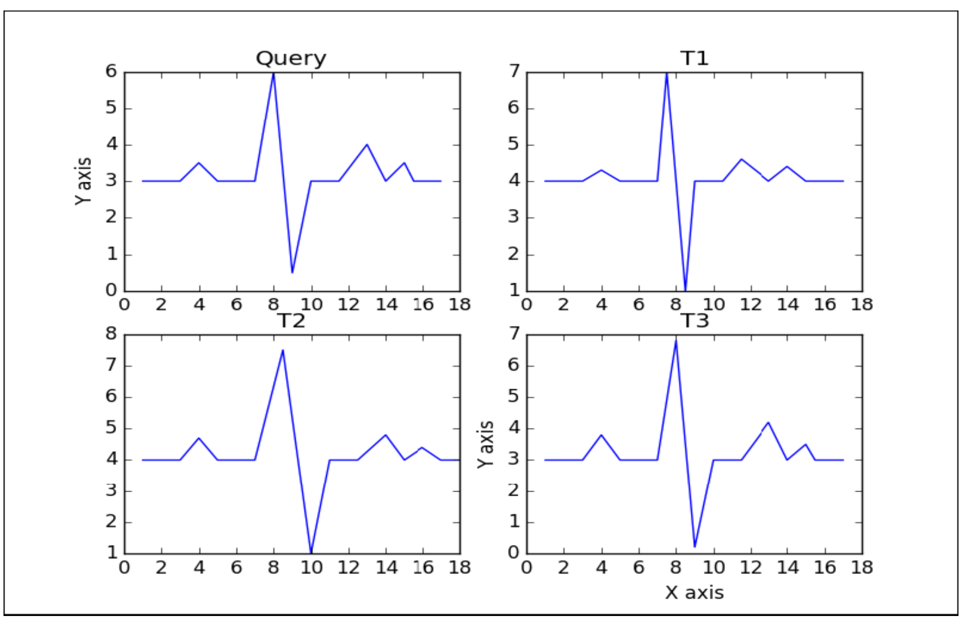

Figure 1. ECG Trajectories - Dataset 1

	(Q,T1)	(Q,T2)	(Q,T3)
DTW	340	304	381
ERP	210	190	250
EDR	12	11	12
LCSS	2	3	2

Table 1. Similarity Distance of Dataset 1

DTW, ERP, EDR and LCSS distance measures are based on the edit distance and these measures are considered to be popular distance measures. These existing edit distance measures are invariant of scaling and translation but not invariant of rotation. Edit distance measures are comparing time series trajectories based on the proximity distance and not based on the shape feature. There is a need to compares time series trajectories based on the shape feature, in order to improve the performance of comparisons. We have provided two motivational examples to discuss the drawback faced by edit distance measures.

Example 1:- Consider the trajectories of synthetic ECG dataset as shown in figure 1. Query trajectory *Q* is compared with the three trajectories for similarity using DTW, ERP, EDR and LCSS measures. The result of comparison of trajectories is shown in table 1. Query trajectory *Q* is similar to the three trajectories T1, T2 and T3. DWT, ERP, EDR and LCSS distance measures are not able to show proper match of the trajectories. This is due to edit distance measures are not comapring trajectories based on the shape feature rather it is simple proximity distance. And also these distance measures are not rotation invariant.

	(Q,T1)	(Q,T2)	(Q,T3)
DTW	361	396	402
ERP	220	206	250
EDR	14	14	13
LCSS	2	1	3

Table 2. Similarity Distance of Dataset 2

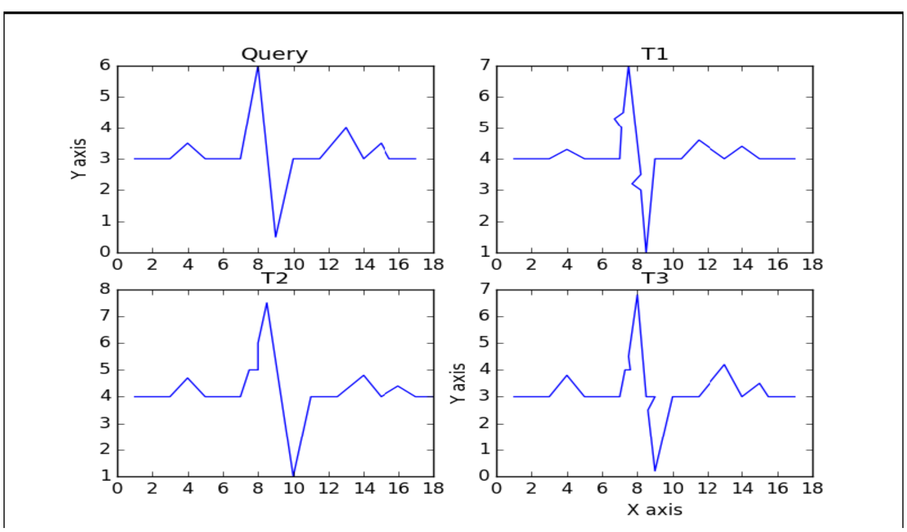

Figure 2. ECG Trajectories with noise - Dataset 2

Example 2: Consider the synthetic ECG dataset 2 of trajectories as shown in figure 2. This synthetic dataset was generated with noise in the trajectories. The query trajectory Q is compared with T1, T2 and T3 trajectories for similarity and result is as shown in table 2. Query trajectory is similar to T1, T2 and T3 trajectories. DWT, ERP, EDR and LCSS distance measures are not able to show proper match of the trajectories. This is due to the existing distance measures are no robust to noise.

We conclude following points from Example 1 and 2 respec-tively. In Example 1, the query trajectory is a proper match with the three trajectories but DTW, ERP, EDR and LCSS distance measures fail to identify as proper match. This is due to fact that, DTW, ERP, EDR and LCSS measures are not comparing trajectories based on shape feature. Also, these edit distance measures are not rotation invariant. So, there is a need to compares trajectories based on shape feature with rotation invariant property.

In Example 2, the query trajectory is similar to the three trajectories from the dataset. This dataset was generated with noise, so that robustness of the edit distance measures are tested. DTW, ERP, EDR and LCSS distance measures fail to identify as proper match. There is a need of a strong distance measure which will be robust to a noise.

In this paper we have proposed Turning Function Based Distance measure in order to tackle the problems faced by DTW, ERP, EDR and LCSS distance measures. Turning Based Distance measure compares trajectories based on shape feature and it is robust to a noise.

Symbols	**Meaning**
R	ECG trajectory $[r_1(x, y), r_2(x, y), r_3(x, y), ..., r_M(x, y)]$
S	ECG trajectory $[s_1(x, y), s_2(x, y), s_3(x, y), ..., s_N(x, y)]$
M	M is the length of R trajectory
N	N is the length of S trajectory and
TFB	Turning Function Based distance measure
Rest(R)	Returned all points of R except first

Table 3. Meaning of symbols used

Contribution of our work in this paper is as follows:

1. We have proposed TFB distance measure to compares ECG trajectories based on the shape feature.

2. TFBdistancemeasuresupportsrotationinvariantproperty.

3. TFB distance measure is robust to a noise.

The rest of paper is organized as follows. In section 2 we present related work. In section 3, we have proposed TurningBasedDistancemeasureforcomparingECGtrajectoriesfor similarity. Section 4, we have discussed TFB algorithm. Section 5, present the experimental results of the proposed TFB techniques. Finally, section 6 concludes the paper.

3. Related Work

Searching of similar time series trajectories is quite interesting problem and quite a few researchers have contributed towards this field. The pioneering work by Agrawal et al. (1993), Agrawal et al. (1995) to compare trajectories for similarity using Euclidean distance measure. The time series are scaled and tranformed eciently and compared for similarity in Kelvin Kam Wing Chu (1999). Berndt and Cliord (1996) introduced DTW to allow a time series trajectories to be stretched to provide better match with another trajectories. The performance of DTW was highlighted in Yasushi Sakurai (2005) and subsequenctly enhanced using lower bounding distance measure with segmentation. Goldin et al. (2004) introduced bounded similarity query with the help of eucledian distance to identify similar time series. The time series are shifted and scaled and then similarity distance is computed , which is then compared with threshold value , to identify similar time series. Yang and Shahabi (2007) have proposed Eros,Muse and Ropes distance measures to identify k-nearest neigbhour using principle component. The distance measures Eros, Muse and Ropes are based on the Eucledian distance measure.

Das et al. (2001) and Vlachos et al. (2002a) applied the LCSS measure to time series trajectory similarity. LCSS measure allow a variable length gap to be inserted during matching of trajectories and hence robust to the noise. In Buzan et al. (2004) extended LCSS measure to compares trajectories of objects from video. First 3D trajectories are extracted from video and then LCSS distance measure is used to compares trajectories. Searching of similar multi dimensional trajectories is explored in Vlachos et al. (2002a). Authors have used LCSS distance measure to compares muti dimensional trajectories. LCSS distance measure is used in conjection with sigmoidal function in Vlachos et al. (2002b) to compares trajectories for similarity. In Vlachos et al. (2006), index multi dimensional trajectories supporting multiple distance measures such as LCSS , DTW are proposed and the index structure was design in a such a way that there was no need to rebuild index again and again. In Keogh et al. (2000) various dimensionality reduction methods were investigated and contributed novel PAA technique to reduce the dimensionality. In Morse and Patel (2007) highlights the extra computation done by LCSS and same is enhanced by fine tuning the threshold value.

Various indexing techniques Faloutsos et al. (1994),Yi et al. (1998) , Chan and chee Fu (1999),Yi and Faloutsos (2000) , Cai and Ng (2004), Keogh et al. (2004a) were proposed to improve the performance of distance measures. Cai and Ng (2004) proposed an eective lower bound technique for indexing. Keogh (2002) have enhanced the indexing method of DTW by modeling exact indexing. Human motions were eciently indexed using bounding rectangle by Keogh et al. (2004b). Rafiei and Mendelzon (2000) introduced technique in which one time index was created and multiple tranformations were applied instead of indexing multiple time , thereby enhancing the performance of similarity search.

Jagadish (1999) proposed landmark similarity distance measure to compares trajectories for similarity. The landmark measure consider the human perception while comparing the trajectories. In Bollobs et al. (1997), the similarity functions is defined using geometric sets , which is deterministic and randomised in nature. In Meratnia and de By (2002), the trajectories were aggregated and compared for the similarities. In Lin and Su (2005), One way distance(OWD) measures were proposed to compare the similarity between trajectories. In Wu et al. (2005), moving objects were detected using Gaussian Hermit moments(GHM) based on the trajectories and also it identify direction of objects. The gestures were recognised based on the trajectories generated by human body in Rubio et al. (2009). In Patwardhan and Roy (2007), gestures modeling and detection of changing shapes were proposed using predictive eigen tracker technique. In Leach et al. (2014), the anomaly behaviour of people in crowded surveillance screnes was identified using statistical aprroach.

Extracting trajectories from videos had been explored by a few researchers. In Buzan et al. (2004) modified LCSS distance measure was used to compares the trajectories. Dyana and Das (2009) proposed new distance measure using gaborfilter to measure the distance between trajectories extracted from videos. Warping distance measure was used in Little and Gu (2001) to compares the trajectories.

Discovering the popular route is non-trivial problem and quite a few researchers have explored popular route problem. In Chen et al. (2011),the popular routes were identified automatically without using road network. First , the newtwork was defined for set of trajectories and then probability function was used to define popularity indicator for the route. The euclidean distance measure was used to calculate the distance between the routes. In Wei et al. (2012), RICK framework was defined to identify popular routes from uncertain trajectories. In Li et al. (2007), the hot route was identified based on the trac density on the road and flowscan framework was developed.

4. Turning Function Based (TFB) distance measure for ECG trajectories

4.1 Basic Terminology Used

• **Definition 1:-** Segment is defined as the line drawn between two points. Let *P1* and *P2* be the two points, then segment is a line joining two points *P1* and *P2*.

• **Definition 2:-** Time Series Trajectory of a moving object is defined as sequence of 2-Dimensional vectors, each describing the position of object at time instants 1 to M . Let *R* and *S* be the two trajectories with length *M* and *N* respectively and are described in 2-Dimensional space as follows:-

$$R = [(t_1, r_1), (t_2, r_2), (t_3, r_3), ..., (t_M, r_M)] \quad (1)$$

$$S = [(t_1, s_1), (t_2, s_2), (t_3, s_3), ..., (t_M, s_M)] \quad (2)$$

• **Definition 3:-** Rest function is defined as a function which return all points of trajectory except the first point.

$$R = [(t_1, r_1), (t_2, r_2), (t_3, r_3), ..., (t_M, r_M)] \quad (3)$$

$$Rest(R) = [((t_2, r_2), (t_3, r_3), ..., (t_M, r_M)] \quad (4)$$

• **Definition 4:-** Turning Function Turning Function [(s)] is defined as the cumulative angle function, which gives the angle between the counterclockwise tangent and the x -axis as a function of the arc length s. θ (s) keeps track of the turning that takes place, increasing with left hand turns, and decreasing with right hand turns. Figure 3 shows Turning Function of a ECG trajectory. The angle between two segments is computed and stored in the turning function. Initially, reference line is drawn and first angle is computed with respect to the reference line. T1 is the first angle with ref line as shown in the figure. Everytime, tangent is drawn with the segment and the angle between tangent and segment is computed. T1 to T13 are the angles of the ECG trajectory in the turning function representation.

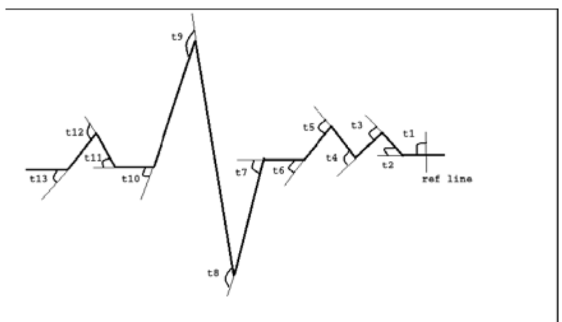

Figure 3. Turning Function

4.2 Turning Function Based Distance measure : TFB

Turning Function is a function which represents ECG time series trajectories with angular distance and segment length. The angular distance is the angle made by segment with the tangent of the other segment. The angle defines a angular shape of the trajectory along with segment length. Thus, angle capture the shape of the trajectories. The angular distance between two turning function is compared and if the dierence of angle between two turning function is less than or equal to threshold value, then the given segment is said to be matched. All the angular distance of turning functions are compared and similarity distance is computed. The similarity distance computed by turning function is compared with the threshold value. If similarity distance is less than or equal to threshold value, then two trajectories are said to be similar.

$$\begin{aligned} TFB(R, S) = min\{&(TFB(Rest(R), Rest(S)) \\ &+ Match(), TFB(Rest(R), S) + 1, \\ &TFB(R, Rest(S) + \mathbf{1}))\} \\ &\text{if } M > 0 \text{ and } N > 0 \end{aligned} \quad (5)$$

The two turning functions are compared using recursive function which calls itself, till terminating condition is reached. When the length of either turning function is equal to 0, then terminating condition is reached and function would returns distance. The minimum value of all possible combination is returned by the recursive function.

Given two trajectories R and S with M and N number of segments in it. Turning Function Based Distance measure compares the two trajectories R and S for similarity by comparing angular distance and segment length. Initially, given trajectory is converted into turning function representation and then turning functions are compared to compute similarity distance between ECG trajectories. Turning Function Based Distance measure is defined as follows:

4.3 Problem Statement

Given Query trajectory $Q = \{(t_1, q_1), (t_2, q_2), ..., (t_M, q_M)\}$, Database trajectories $D = \{(d_1, d_2, d_3, ..., d_{NT})\}$ where $\{d_1, d_2, d_3, ..., d_{NT}\}$ are database trajectories and a user specified threshold δ. The query trajectory is compared with database trajectories for similarity. The two trajectories Q and S are said to be similar if and only if similarity distance returned by TFB(Q,S) measure is less or equal to where Q is query trajectory and S is belong to D.

4.4 Properties of TFB measure

Our proposed distance measure supports various properties and these properties are useful for making TFB distance measure stronger. Following are the properties supported by TFB distance measure.

- **Property 1:-** TFB measure always returns positive value.

$$TFB(R, S) \geq 0$$

- **Property 2:-** TFB measure is symmetric in nature.

$$TFB\,(R, S) = TFB\,(S, R)$$

- **Property 3:-** TFB measure supports Triangular Inequality property.

$$TFB(R; P) + TFB(P; S) \geq TFB(R; S)$$

- **Property 4:-** TFB measure supports metric property since it supports triangular inequality, positiveness and symmetric properties.

Algorithm 1: TFB Function

Input: R and S : Input Trajectories

Output: Similarity Distance : SimDist

if *Length(R) ==0 or Length(S)==0* **then**
    SimDist= 1
    return SimDist
else
    SimDist= min fTFB(Rest(R), Rest(S)) + Match() ,
    TFB(Rest(R),S) + 1 ,TFB(R,Rest(S)) +1 g
return *S imDist*

Algorithm 2: TurningFunc Function

Input: T : Input trajectory

Output: T : Turning Function

for *each segment* $S \in T$ **do**
    T’.length=Length(S)
    T’.angle = Angle(S)

return *T′*

Algorithm 3: TFB Algorithm

Input: Q : Query Trajectory, D : Raw Database Trajectories

Output: ST : Similar Trajectories

```
1 Q'=TurningFun(Q)
2 for each point S ∈ D'_i do
3 |    S'=TurningFunc(S)
4 |    SimDist = TFB(Q',S')
5 |    if SimDist ≤ δ then
6 |    |    S is similar trajectory to Q
7 |_   |_   Append S to ST
8 return S T
```

5. TFB Algorithm

TFB algorithm accepts Query trajectory *Q* and Database trajectories *D*. The query trajectory *Q* is compared with Database trajectories D to identify all similar trajectories as that of query. The similarity distance of query trajectory and database trajectories are computed and if the similarity distance is less or equal to threshold value, then such trajectories are returned from TFB algorithm.

Line 1 of TFB algorithm, makes a call to TurningFunc with Q as an argument. TurningFunc converts Q trajectory into turning function representation. Line 2, iterates through all the trajectories of the database. Line 3, makes a call to TurningFunc with S as an argument. Turning function converts database trajectory S into turning function representation. Line 4, makes a call to TFB recursive function with Q' and S' as an argument. The similarity distance is returned from TFB function. Line 5, checks if similarity distance is less than or equal to threshold value, then S' is similar to query trajectory and it is stored in the ST.

Turningfunc algorithm accepts a trajectory and convert into turning function representation. Line 1, iterates through all the segments of the trajectory. Line 2, computes length of each segment. Line 3, computes angle between two consecutive segments. T' trajectory is returned from the algorithm.

TFB function accepts two turning functions as an argument. Line 1, check the length of the two trajectories and if length of either trajectory is zero, then return 1. Line 5, makes a recursive calls with three arguments. The minimum of the three arguments is returned as the similarity distance value. Match function compares the angular distance of the current segments. If the angular distance is within the permissible limit, then zero value is returned else 1 value is returned.

TFB algorithm accept Query and Database trajectories as an input and it returns similar trajectories as an output. Let M and N be the number of segments in R and S trajectories respectively. TurningFunc function converts raw trajectories into turning function representation. If there are NT number of the trajectories, then time complexity of the function is given as O(N*NT) where N is the number of segments of trajectory and NT is the number of trajectories. TFB function is a recursive function which can be implemented in dynamic programing with a complexity of O(N*M) where N and M are length of the trajectories. Thus, total time complexity of TFB distance measure to search similar trajectories is O(N*M*NT).

6. Experimental Study

6.1 System Configuration

Experimental study was carried out on Pentium V processor with 4GB of RAM and 500GB of hard disk memory. All the programs were successfully implemented using C++ language. The g++ compiler was used to compile the C++ programs. Ubuntu 12.04 operating system was used to carry out experimental study.

6.2 Characteristic of Time Series Datasets

We have used real time and synthetic time series datasets for our experimental study.

1. Arrhythmia Data Set (ECG)Lichman (2013):- This database contains 279 attributes, 206 of which are linear valued and the rest are nominal. Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.

2. EEG Database Data SetLichman (2013):- This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on subject's scalps which were sampled at 256 Hz (3.9-msec epoch) for 1 second. There were two groups of subjects: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set. When two stimuli were shown, they were presented in either a matched condition where S1 was identical to S2 or in a non-matched condition where S1 diered from S2.

3. Heart Disease Data SetLichman (2013):- This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

4. Synthetic ECG dataset :- This is a synthetic dataset contains 20,000 trajectories of 1000 patients. The program was written to simulate 1000 heart patients behavior and 20 ECGs were generated for each patient.

6.3 Results and Interpretations

Initally, we have carried out experiment with small datasets consisting of a few trajectories. This is to make sure that our proposed distance measure and existing distance measures are producing correct output. The query trajectory is compared with three trajectories from the dataset. There are two types of dataset used for the experiment. Dataset 1 is prepared without introducing noise in the trajectories. The second dataset, Dataset 2 is prepared by inserting noise in the trajectories. Dataset 2 is mainly used to check the robustness of the distance measures with noise.

DTW, ERP, EDR, LCSS and TFB distance measures were applied to the ECG dataset of 1 and their results are shown in table 4. DTW, ERP, EDR, LCSS distance measures fail to identify as proper match even though query trajectory is similar to *T1*,*T2* and *T3* trajectories. DWT, ERP distance measures similarity distance is high indicating dissimilarity between trajectories. Where as EDR distance measure is showing high value with respect to the length of the trajectories indicating that there is mismatch of the trajectories. LCSS distance measure is showing low value indicating that there is mismatch of the trajectories. The results from the table reveal that TFB measure is able to match query trajectory *Q* with *T1*, *T2* and *T3* trajectories. TFB distance measure is able to identifies similar trajectories since TFB distance measure is shape based measure.

	(Q,T1)	**(Q,T2)**	**(Q,T3)**
DTW	340	304	381
ERP	210	190	250
EDR	12	11	12
LCSS	2	3	2
TFB	1	1	1

Table 4. Similarity Distances on Dataset1

DTW, ERP, EDR, LCSS and TFB distance measures were applied to ECG dataset 2 and results are shown in table 5. This datasets was generated with noise being introduced in the dataset. DTW, ERP, EDR, LCSS distance measures fail to identify as proper match even though query trajectory is similar to *T1*,*T2* and *T3* trajectories. DWT, ERP distance measures similarity distance is high indicating dissimilarity between trajectories. Whereas EDR distance measure is showing high value with respect to the length of the trajectories indicating that there is mismatch of the trajectories. LCSS distance measure is showing low value indicating that there is mismatch of the trajectories. The results from the table reveals that TFB measure is able to match query trajectory *Q* with *T1, T2* and *T3* trajectories. TFB distance measure is able to match the trajectories since TFB distance measure is shape based and it is robust to the noise.

	(Q,T1)	(Q,T2)	(Q,T3)
DTW	361	396	402
ERP	220	206	250
EDR	14	14	13
LCSS	2	1	3
TFB	2	1	2

Table 5. Similarity Distances on Dataset 2

To compute accuracy of the distance measures, we have used 30 correct samples from each datasets. We have applied the edit distance measures to dierent datasets and recorded the total correct match found out of 30 correct samples. Then the accuracy of edit distance measure is defined as the total number of correct samples identified using edit distance measure divided by the total number of correct samples.

Distance Measure	Arrhythmia Data Set (ECG)	Accuracy
DTW (30)	14	46
ERP (30)	14	46
EDR (30)	14	46
LCSS (30)	13	43
TFB (30)	18	60

Table 6. Comparison of TFB with other distance measures

Distance Measure	Arrhythmia Data Set (ECG)	Accuracy
DTW (30)	11	36
ERP (30)	12	40
EDR (30)	11	36
LCSS (30)	11	36
TFB (30)	15	50

Table 7. Comparison of TFB with other distance measures

Table 6 shows the accuracy of DTW, ERP, EDR, LCSS and TFB distance measures. The accuracy of DTW measure was 46, ERP was 46, EDR was 46 and LCSS was 43. These distance measures show low accuracy since they are not based on shape and do not support rotation invariant property. The accuracy of TFB distance measure was 60. The accuracy of TFB distance measure is at higher side since it compares trajectories based on shape feature and it is rotation invariant.

Table 7 shows the accuracy of DTW, ERP, EDR, LCSS and TFB distance measures with noisy dataset. The accuracy of DTW measure was 36, ERP measure was 40, EDR measure 36 and LCSS was 36. These distance measures show low accuracy since they are not robust to noise. The accuracy of TFB distance measure was 50. The accuracy of TFB distance measure is at higher side since it is robust to noise.

Execution time of dierent edit distance measures were recorded. Execution time is the time taken by the distance measure to generate the output. Execution time of dierent distance measures was recorded for dierent datasets. The purpose of this experiment was to identify the performance of the edit distance measure with dierent datasets.

The graph in figure 4, shows the performance of the DTW, ERP and TFB distance measures. Since the performance of edit distance measures are almost same with respect to shape based comparison, we have carried out experimental study with DTW and ERP distance measures. Other edit distance measures can be implemented in similar line and their performance can be analyzed. The Four datasets were used to carry out experimental study such as Arrhythmia Data Set (ECG), EEG Database Data Set, Heart Disease Data Set and Synthetic ECG dataset. The execution time of DTW and ERP distance measures were almost same, where as TFB has some extra overhead due to extra processing of turning function representation.

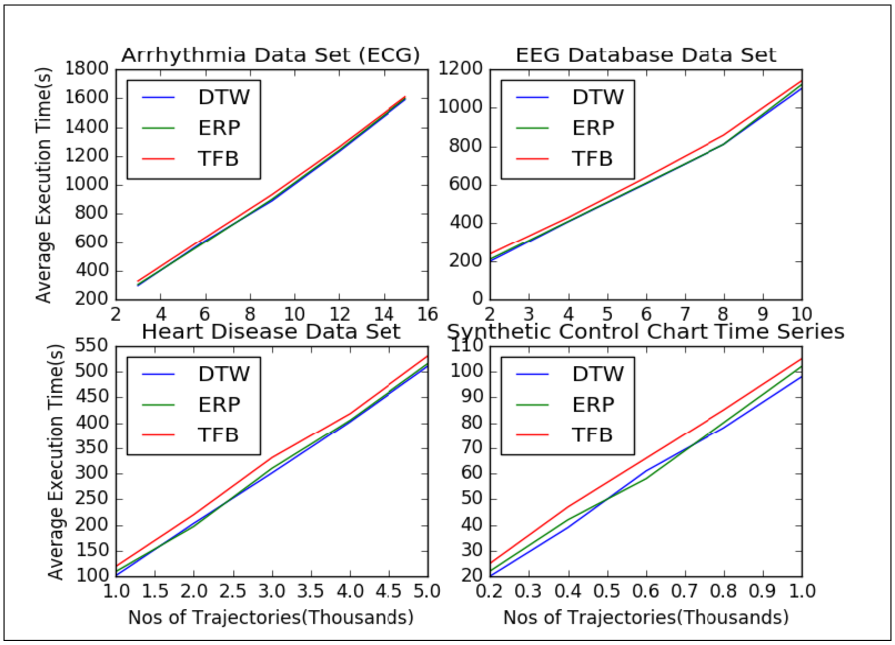

Figure 4. Performance of DTW, ERP and TFB

7. Conclusion

We have proposed Turning Function Based Distance measure to compares ECG time series trajectories for similarity. TFB distance measure compares trajectories based on the shape feature. TFB distance measure supports scaling, translation, rotation invariant property and robust to noise. Experimental study was carried out on real time and synthetic datasets of ECG trajectories. Our experimental results show that TFB measure compares the trajectories based on shape of the trajectories and robust to the noise. The average accuracy of DTW, ERP , EDR , LCSS distance measures without noise was 46 and with noise was 38. The average accuracy of TFB distance measure without noise was 60 and with noise was 50. Thus, TFB distance measure is ecient compared to DTW, ERP, EDR, LCSS distance measures.

References

[1] Agrawal, R., Faloutsos, C., Swami, A. (1993). Ecient Similarity Search, *In:* Sequence Databases, *In:* : FODO, Springer Verlag, 69–84

[2] Agrawal, R., ip Lin, K., Sawhney, H. S., Shim, K. (1995). Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases, *In:* : In-VLDB, 490–501.

[3] Kelvin Kam Wing Chu, M. H.W. (1999). Fast Time Series Searching with Scaling and Shifting, *In:* ACM-PODS, 237–248.

[4] Berndt, D. J., Cliord, J. (1996). Finding Patterns in Time Series: A Dynamic Programming Approach, *In:* Advances in Knowledge Discovery and Data Mining, 229–248.

[5] Yasushi Sakurai, C. F., Masatoshi Yoshikawa. (2005). Fast Similarity Search under the Time Warping Distance, *In:* ACM-PODS, 326–337.

[6] Goldin, D. Q., Millstein, T. D., Kutlu, A. (2004). Bounded similarity querying for time-series data, Information and Computation 194 (2) 203 – 241, URL http://www.sciencedirect.com/science/article/pii/S0890540104001245, special Issue Commemorating the 50th Birthday Anniversary of Paris C. Kanellakis.

[7] Yang, K., Shahabi, C. (2005). An ecient k nearest neighbor search for multivariate time series, Information and Computation 205 (1) (2007) 65 – 98, URL http://www.sciencedirect.com/science/article/pii/S0890540106001039, special Issue: fTIMEg

[8] Das, G., Gunopulos, D., Mannila, H. (2001). Time-Series Similarity Problems andWell-Separated Geometric Sets, *Nord. J. Comput.* 8 (4) 409–423, URL http://www.cs.helsinki.fi/njc/References/bollobas2001:409.html.

[9] Vlachos, M., Kollios, G.., Gunopulos, D. (2002). Discovering Similar Multidimensional Trajectories, *In:* ICDE, 673–684, a.

[10] Buzan, D., Sclaro, S., Kollios, G. (2004). Extraction and Clustering of Motion Trajectories in Video, *In:* : 17th InternationalConference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004., 521–524.

[11] Vlachos, M., Gunopulos, D., Kollios, G. (2002). Robust Similarity Measures for Mobile Object Trajectories., *In:* : DEXA Workshops, IEEE Computer Society, 721–728.

[12] Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E. J. (2006). Indexing MultiDimensional TimeSeries with Support for Multiple Distance Measures, VLDB J. 15 (1) 1–20.

[13] Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. (2000). Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases, *JOURNAL OF KNOWLEDGE AND INFORMATION SYSTEMS* 3 263–286.

[14] Morse, M. D., Patel, J. M. (2007). An ecient and accurate method for evaluating time series similarity., *In:* : C. Y. Chan, B. C. Ooi, A. Zhou (Eds.), SIGMOD Conference, ACM.

[15] Faloutsos, C., Ranganathan, M., Manolopoulos, Y. (1994). Fast Subsequence Matching in Time-Series Databases

[16] Yi, B.-K., Jagadish, H. V., Faloutsos, C. (1998). Ecient Retrieval of Similar Time Sequences Under Time Warping, *In:* S. D. Urban, E. Bertino (Eds.), Proceedings of the Fourteenth International Conference on Data Engineering, Orlando, Florida, USA, February 23-27, 1998, 201–208.

[17] Chan, K.-P., chee Fu, A.W. (1999). Ecient Time Series Matching by Wavelets, *In:* ICDE, 126–133

[18] Yi, B.-K., Faloutsos, C. (2000). Fast Time Sequence Indexing for Arbitrary Lp Norms.,*In:* A. E. Abbadi, M. L. Brodie, S. Chakravarthy, U. Dayal, N. Kamel, G. Schlageter, K.-Y. Whang (Eds.), VLDB, Morgan Kaufmann. URL http://dblp.uni-trier.de/db/conf/vldb/vldb2000.html#YiF00

[19] Cai, Y., Ng, R. (2004). Indexing Spatio-temporal Trajectories with Chebyshev Polynomials, *In:* Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, SIGMOD '04, ACM, New York, NY, USA, URL http://doi.acm.org/10.1145/1007568.1007636

[20] Keogh, E. J., Palpanas, T., Zordan, V. B., Gunopulos, D., Cardle, M. Indexing Large Human-Motion Databases., in: M. A. Nascimento, M. T. zsu, D. Kossmann, R. J. Miller, J. A. Blakeley, K. B. Schiefer (Eds.), VLDB, Morgan Kaufmann, , URL http://dblp.uni-trier.de/db/conf/vldb/vldb2004.html# PalpanasCGKZ04, 2004a.

[21] Keogh, E. (2002). Exact Indexing of Dynamic Time Warping

[22] Keogh, E., Palpanas, T., Zordan, V. B., Gunopulos, D., Cardle, M. (2004). Indexing Large Human-motion Databases, in: Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, VLDB '04, VLDB Endowment, URL http://dl.acm.org/ citation.cfm?id=1316689.1316757, b.

[23] Rafiei, D., Mendelzon, A. O. Querying Time Series Data Based on Similarity, *IEEE Transactions on Knowledge and Data Engineering* 12 (2000) 675–693.

[24] Jagadish, H. V. Review - Landmarks: a New Model for Similarity-based Pattern Querying in Time Series Databases, *ACM SIGMOD Digital Review* 1.

[25] Bollobs, B., Das, G., Gunopulos, D., Mannila, H. (1997). Time-Series Similarity Problems and Well-Separated Geometric Sets.,

In: Symposium on Computational Geometry, 454–456.

[26] Meratnia, N., de By, R. A. (2002). Aggregation and comparison of trajectories., in: A. Voisard, S.-C. Chen (Eds.), ACM-GIS, ACM.

[27] Lin, B., Su, J. (2005). Shapes based trajectory queries for moving objects., in: C. Shahabi, O. Boucelma (Eds.), GIS, ACM.

[28] Wu, Y., Shen, J., Dai, M. (2005). Trac object detections and its action analysis., *Pattern Recognition Letters* 26 (13) 1963–1984.

[29] Rubio, J. P. B., Marfil, R., Bandera, A., Rodrguez, J. A., Molina-Tanco, L. , Hernndez, F. S. (2009). Fast gesture recognition based on a two-level representation., *Pattern Recognition Letters* 30 (13) 1181–1189.

[30] Patwardhan, K., Roy, S. D. (2007). Hand gesture modelling and recognition involving changing shapes and trajectories, using a Predictive EigenTracker., *Pattern Recognition Letters* 28 (3) 329–334.

[31] Leach, M. J. V., Sparks, E. P., Robertson, N. M. (2014). Contextual anomaly detection in crowded surveillance scenes, *Pattern Recognition Letters* 44 71–79.

[32] Dyana, A., Das, S. (2009). Trajectory representation using Gabor features for motionbased video retrieval, *Pattern Recognition Letters* 30 (10) 877–892.

[33] Little, J. J., Gu, Z. (2001). Video retrieval by spatial and temporal structure of trajectories., *In*: M. M. Yeung, C.-S. Li, R. Lienhart (Eds.), Storage and Retrieval for Media Databases, vol. 4315 of SPIE Proceedings, SPIE, 545–552.

[34] Chen, Z., Shen, H. T., Zhou, X. (2011). Discovering popular routes from trajectories., *In*: S. Abiteboul, K. Bhm, C. Koch, K.-L. Tan (Eds.), ICDE, IEEE Computer Society.

[35] Wei, L.-Y., Zheng, Y., Peng, W.-C. (2012). Constructing popular routes from uncertain trajectories., *In*: Q. Y. 0001, D. Agarwal, J. Pei (Eds.), KDD, ACM.

[36] Li, X., Han, J., Lee, J.-G., Gonzalez, H. (2007). Trac Density-based Discovery of Hot Routes in Road Networks, *In*: Proceedings of the 10th International Conference on Advances in Spatial and Temporal Databases, SSTD'07, Springer- Verlag, Berlin, Heidelberg, 441–459.

[37] Lichman, M. (2013). UCI Machine Learning Repository, http://archive. ics.uci.edu/ml