P-Wave Onset Point Detection for Seismic Signal Using Bhattacharyya Distance

Bikash Chandra sahana *sahana@nitp.ac.in*
Department of Electronics & Communication Engineering
National Institute of Technology Patna
Patna, 800005, India

M.M Choudhary *mmc@nitp.ac.in*
Department of Electrical Engineering
National Institute of Technology Patna
Patna, 800005, India

Abstract

In seismology Primary p-wave arrival identification is a fundamental problem for the geologist worldwide. Several numbers of algorithms that deal with p-wave onset detection and identification have already been proposed. Accurate p- wave picking is required for earthquake early warning system and determination of epicenter location etc. In this paper we have proposed a novel algorithm for p-wave detection using Bhattacharyya distance for seismic signals. In our study we have taken 50 numbers of real seismic signals (generated by earthquake) recorded by K-NET (Kyoshin network), Japan. Our results show maximum standard deviation of 1.76 sample from true picks which gives better accuracy with respect to ratio test method.

Keywords: P-Wave Picking, Seisming Signal Processing, Bhattacharyya Distance, Seismic Data.

1. INTRODUCTION

In seismology from the seismologist point of view it is extremely important to detect and accurately prediction of the first P-wave arrival. The p-wave arrival tells us important information related to event detection, identification, source acquisition of triggered seismic data, mechanism analysis, etc. In old days this work have been accomplished manually in a visual way. But in the era of information and communication technology ,it can be done by computer programs.

In literature various methods have been implemented for automatic p- phase onset arrivals. The first idea and attempts towards automatic picking was proposed base on the ratio of a Short Term Average (STA) and Long Term Average (LTA) of some Characteristic Function (CF) of the real seismic data. Basically in the noisy area the ration should remain substantially constant. When the signal emerges the STA should be able to capture the change most quickly than the LTA, resulting a sudden rise of the ratio values. By comparison of the STA/LTA ratio with predetermined threshold p onset can be determined. The Characteristic function (CF) is proposed by Allen [1], [2] is given as weighted sum of the squared amplitude and the squared derivative of the signal. Earle and Shearer [3] proposed a method where he used a function that defines envelope of the seismogram. Envelope Function is obtained by the squared root of the sum of the squared values and the squared Hilbert transform of the seismogram. By considering seismogram as composed of two different stationary processes that divides at onset point. The problem is modeled into AR model of seismic data [4], [5]. AR model based picker requires large SNR.

Anant and Dowla [6] applied the Discrete Wavelet Transform (DWT) and used polarization and amplitude information contained in the wavelet. Gendron et al [7] extract feature based on wavelet coefficients of seismic data and then jointly detected and classified seismic events via Bayes theorem. Zhang et.al [8] did the soft thresholding on the DWT coefficient to denoise the data the calculate Akaike Information Criterion (AIC) like sequence without fitting AR models.Der and Shumay [9], used modifies version of the CUMSUM algorithm for the detection of the multiple variance changes in time series. Nakamula et. al. [10] divide a record into equal length frames and check the local and weak stationary of each interval using the theory of the Langevin equations. They took assumption that the process is stationary only if background noise is present. The stationary state will break abruptly when seismic signal arrives and the frame include both background noise and samples of the p-wave.

Hafez and Kohda [11], used undecimated version of wavelet transfor i.e. Maximal overlap discrete wavelet transform (MODWT) for manually determining clear P-wave arrivals of weak events. They determine MODWT coefficients of seismic events calculated at different scales after that detailed feature at different scale are determined. An adaptive thresholding method is used. Pikoulis and Psarakis [12], used a nonlinear transformation based on the notion of the length of seismogram and a ration test. First they determine distance between adjacent samples and then applying ration ratio test on this. The maxima is obtained at the onset point further he modifies the maxima determination problem to estimate the onset.

Remaining part of the paper is organized as follow. In the section 2 problem is described. In the section 3 a nonlinear transformation of the real raw seismic data based on the length of seismogram is defined and a Bhattacharyya distance calculation method criterion is used for further detection and identification of p-wave onset. In the section 4 our experimental results are presented. Finally section 5 contains our conclusion.

2. SEISMIC DATA CHARACTERIZATION

Seismic data record consists of seismic events and seismic noise. If y_n , $n = 0,1,\ldots,$N denotes seismic event datasamples recorded from given station, and Let us assume that during recording interval, K seismic events occurred. If we denote $s_n^k, n = 0,1,\ldots,N_k$ the signal produced by k-th seismic event (earthquake) and the $n_k$the corresponding arrival time, then the combined event can be expressed as;

$$y_n = \sum_{k=1}^{K} s_{n-n_k}^k + e1_n + e2_n + e3_n \qquad (1)$$

Where $e1_n, e2_n, e3_n$ are seismic noises. Seismic noise $e1_n$ is caused by traffic, wind, and machinery which have period of range 0.1 seconds. Seismic noise $e2_n$ have period between 0.1 to 2.0 seconds and occurs in town areas. Seismic noise $e3_n$have period of 3 to 10 seconds caused by storm over oceans. Seismologist preprocessed the data by bandpass filtering operation [14]. In proposed algorithm it is assumed that Seismic event start at some time n1 and end at n2. Earthquake have higher power than the noise $e3_n$. The time duration of seismic noise segment, $e1_n, and\ e2_n$ are shorter than that of earthquake. The frequency content of a seismic event y_n much higher than the seismic noise $e3_n$.

3. PROPOSED ALGORITHM USING BHATTACHARYA DISTANCE

Flowchart of our Proposed algorithm is shown in figure (7) corresponding used terms are explained below.Let us consider a discrete time signal y_n obtain from the sampling of its continuous time component $y(t)$ with a sampling period of T_s, so that $y_n = y(nT_s)$. Let us also define $\Delta\mathcal{L}_n, n = 1,2,\ldots$ as the Euclidean distance of the line segment connection consecutive pairs of the points $((n-1)T_s, y_{n-1})$ and $((n)T_s, y_n)$,i.e.;

$$\Delta\mathcal{L}_n = ((y_n - y_{n-1})^2 + T_s^2)^{\frac{1}{2}} = T_s(\frac{y_n - y_{n-1})}{T_s} + 1)^{\frac{1}{2}} \qquad (2)$$

In order to give more physical interpretation in the above defined quantity. Let us assume the case of noiseless, i.e. $e1_n, e2_n, and\ e3_n$ are zeroes. If we consider that the first order backward differences of signal y_n appeared in equ. (2) constitute and approximation of the derivative $\dot{y}(t)$ of function $y(t)$ at the sampling points nT_s, i.e.

$$\frac{y_n - y_{n-1})}{T_s} \approx \dot{y}(t)|_{t=nT_s}, \qquad (3)$$

Then $\Delta\mathcal{L}_n/T_s$, can be considered as the approximation of the instantaneous change of the length of the curve (let us denote it by $\mathcal{C}(y)$), defined by the following relation:

$$\mathcal{L}(t) = (y^2(t) + 1)^{\frac{1}{2}}, \qquad (4)$$

From Equation. (2) and (3) we get:

$$\frac{\Delta\mathcal{L}_n}{T_s} \approx \mathcal{L}(t)|_{t=nT_s}, \qquad (5)$$

It has to be noted that equation (4) represent the instantaneous change in the length of $\mathcal{C}(y)$), as a function of the first derivative of $y(t)$. And also noted that the equation (2) is obtained after highly nonlinear filtering operation on the original signal with high frequency is enhanced and low frequency is suppressed at the same time.

However, when noise is present in the seismic signal y_n we can model it in terms of stochastic process. That means $\Delta\mathcal{L}_n$ are considered as random variables (RVs). Consider a real seismic data shown in figure (1) and corresponding evaluation of $\Delta\mathcal{L}_n$ is shown in figure (2). and corresponding evaluation of $\Delta\mathcal{L}_n$ is shown in figure (2).

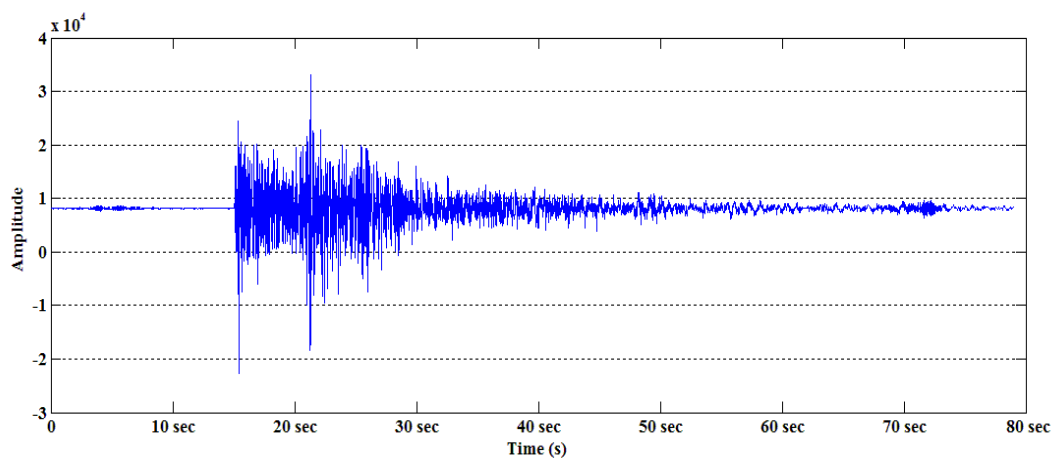

FIGURE 1: Real Raw Seismic Data (Amplitude vs Time) .

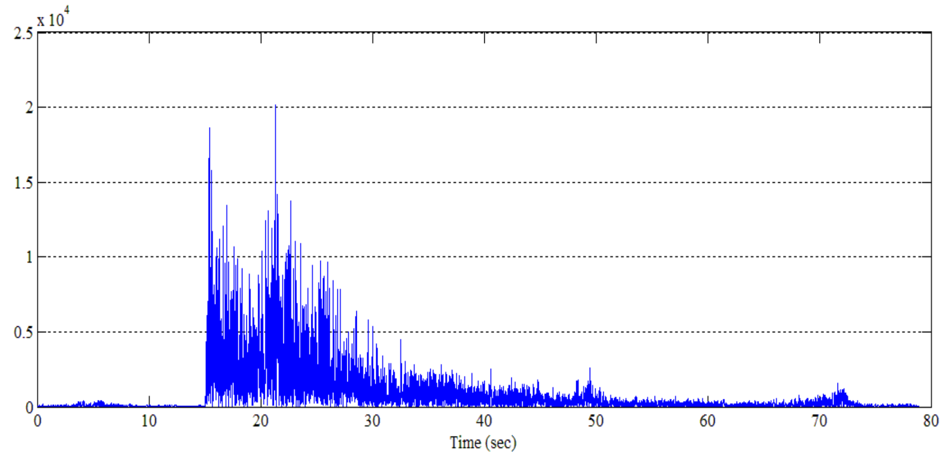

FIGURE 2: Value of $\Delta\mathcal{L}_n$ of Seismogram ($\Delta\mathcal{L}_n$ vs time).

Now we use our statistic measure Bhattacharyya distance for our problem. Let us consider two sequences of $\Delta\mathcal{L}_n$ as shown below.

$$\nabla\mathcal{L}_n^{N+} = [\Delta\mathcal{L}_n, \Delta\mathcal{L}_{n+1}, \dots, \Delta\mathcal{L}_{n+N-1}] \qquad (6)$$

$$\nabla\mathcal{L}_n^{M-} = [\Delta\mathcal{L}_{n-M+1}, \Delta\mathcal{L}_{n-M}, \dots, \Delta\mathcal{L}_{n-1}] \qquad (7)$$

Where $\nabla\mathcal{L}_n^{N+}$ forward window of length of size N and $\nabla\mathcal{L}_n^{M-}$ backward window of length M. We assumed that $y(t)$ and its derivative $\dot{y}(t)$ are both continuous in $\mathbb{R}$. Now assuming these two sequences are normally distributed. Let us consider windows $\nabla\mathcal{L}_n^{N+}$ and $\nabla\mathcal{L}_n^{M-}$ represented by class $n,1$ and class $n,2$ respectively, Then Bhattacharyya distance [14] between these two classes can be determined by the following relation.

$$b_n = \frac{1}{8}(\mu_{n,1} - \mu_{n,2})^T \left[\frac{\sum_{n,1} + \sum_{n,2}}{2}\right]^{-1} + \frac{1}{2} ln \frac{|(\sum_{n1} + \sum_{n2})/2|}{|\sum_{n,1}|^{1/2} |\sum_{n,1}|^{1/2}} \qquad (8)$$

For $n = 0,1,2,\dots$

Where $\mu_{n,1}$ and $\sum_{n,1}$ are the mean vector and covariance matrix of class $n,1$. And $\mu_{n,2}$ and $\sum_{n,2}$ are the mean vector and covariance matrix of class $n,2$. The distance constitutes the proposed test statistic for the problem. The evaluation of the b_n is shown below in figure (3).

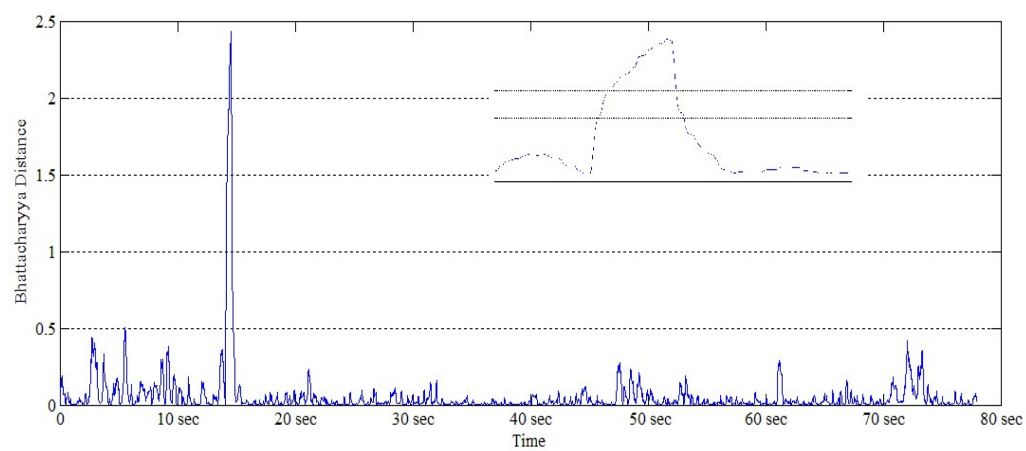

FIGURE 3: Value of b_n Obtained from $\Delta\mathcal{L}_n$ sequence . (bn vs Time).

b_n is calculated with window size $M = N = 40$, in figure (3) a zoomed portion of the plot focusing on the main globe, is displayed in the upper right corner of the same figure.

Let us now concentrate on the solution of the desired problem, namely on the estimation of n_0. A natural selection to achieve our goal would be the solution of the following maximization problem;

$$\hat{n}_0 = arg \max_n b_n \, . \qquad (9)$$

the location where b_n attain its maximum value,

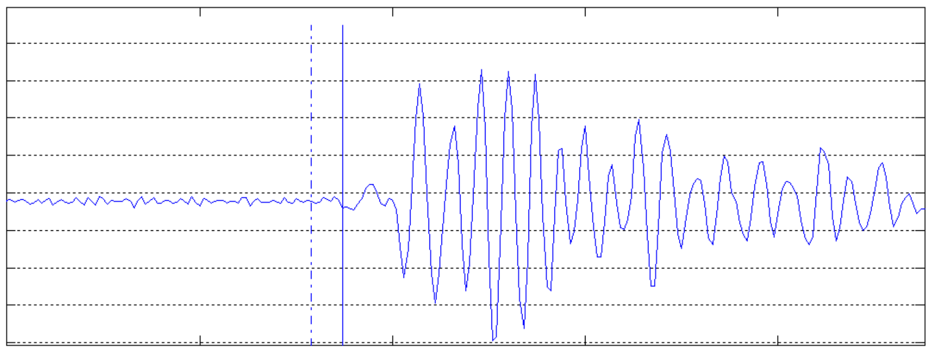

FIGURE 4: Estimation of n_0 from the value of b_n (bn vs time n).

Estimation of time instance where P wave starting point (n0) in equation (9) is shown above in figure (4) on real seismic data after evaluating b_n shown below.

In figure (4) the solid vertical line correspond to true value of n_0 and the dashed vertical line shows the estimated value of n_0 .

4. RESULT & DISCUSSION

We did our simulation work on software package MATLAB. And all results are generated in MATLAB environment. The proposed scheme is tested on 50 real seismogram each containing one event that is recorded with sampling rate 100 sample per second. The proposed algorithm is used on these data and then error is calculated with respect to manual picks. Ratio test algorithm [12] is also applied on these seismic recorded data and further error is calculated. The performance of seismic pick detection of proposed and ratio test method is shown in figure (6) and (7) with histogram representation correspondingly. Our method gives better resulted than ratio test method. In our result the histogram is concentrated toward true p-wave onset.

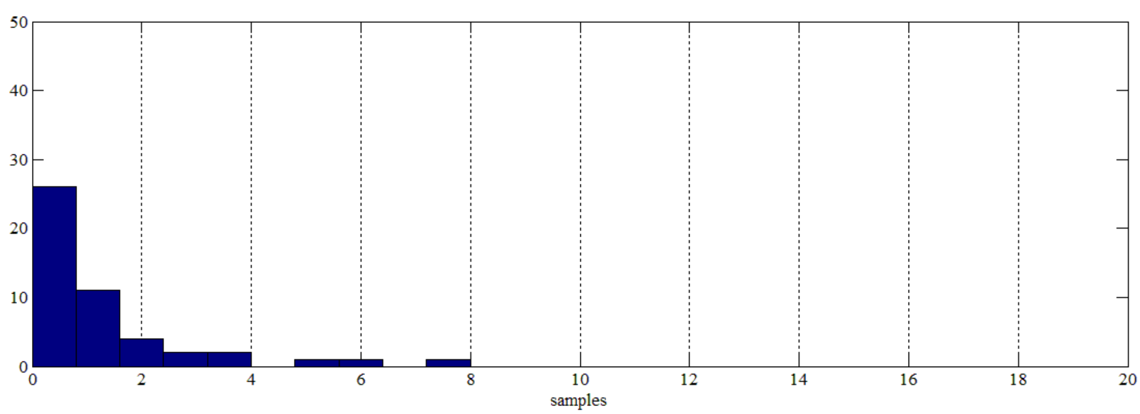

FIGURE 5: Histogram of picking errors obtained by application of proposed methods.

In the proposed process, It applies highly nonlinear filtering operation on the original signal with high frequency is enhanced and low frequency is suppressed at the same time. some novel statistical technique may be applied to reduce the error in great extend in future as extension to the present work.

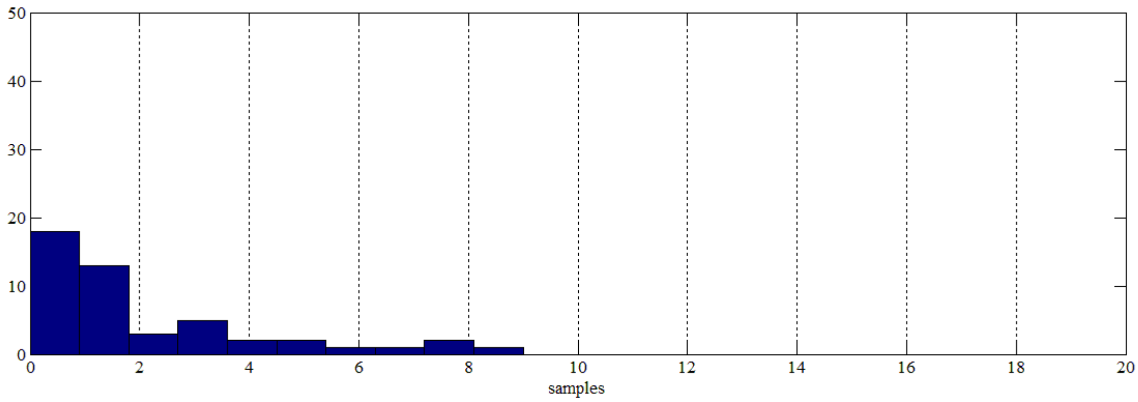

FIGURE 6: Histogram of picking errors obtained by application of ratio test method.

5. CONCLUSION

In seismic signal processing it is very important to pick P- wave onset point. Here our algorithm shows very promising result . Our results show maximum standard deviation of 1.76 sample. from true picks which gives better accuracy with respect to existing ratio test method method. In this paper ,Comparison has been done in performance of P-wave onset point detection between proposed method with ratio test method .Error in P-wave onset point detection is basically difference between point of actual P wave arrival point ,analyzed manually in the time(n) axis with calculated value of P-wave arrival point using proposed algorithm. All existing methods like ratio test methods are porn to error in terms of different sample numbers but our method shows less error and maximum standard deviation in errors in terms of samples are also less which results better accuracy in P-wave onset point detection.

6. REFERENCES

[1] R.V. Allen.(1978). “Automatic earthquake recognition and timing from single traces”, Bull. Seism. Soc. Am. 68, 1521-1532, [1978].

[2] R.V. Allen.(1982). “Automatic phase pickers: their present and future prospects”, Bull. Seism. Soc. Am. 68, S225-S242, [1982]

[3] P. Earle, and P. Shearer.(1994). “Characterization of global seismograms using an automatic picking algorithm”, Bull. Seism. Soc. Am. 84, no. 2, 366-376, [1994].

[4] A. Kushnir, V. Lapshin, V. Pinsky, and J. Fyen.(1990). “Statistically optimal event detection using small array data”, Bull. Seism. Soc. Am. 80, no. 6b, 1934-1950, [1990].

[5] T. Takanami, and G. Kitagawa.(1988) “A new efficient procedure for the esti- mation of onset times of seismic waves” Journal of Phys. Earth 36, 267-290, [1988].

[6] S.K Anant, and F.U. Dowla.(1997) “Wavelet transform methods for phase identification in three-component seismogram”, Bull. Seism. Soc. Am. 87, 1598-1612, [1997].

[7] P. Gendron, J. Ebel, and D. Manolakis.(2000). “Rapid joint detection and classification with wavelet bases via Bayes theorem”, Bull. Seism. Soc. Am. 90, 764-774, [2000]

[8] H. Zhang, C. Thurber, and C. Rowe.(2003). “Automatic P-Wave Arrival Detec- tion and Picking with Multiscale Wavelet Analysis for Single-Component Recordings”, Bull. Seism. Soc. Am. 93, no. 5, 1904-1912, [2003].

[9] Z.A. Der, and R.H. Shumway.(1999). “Phase onset time estimation at regional distances using the CUSUM algorithm”, Phys. Earth Planet. Int. 113, no. 2, 227-246, [1999].

[10] S. Nakamula, M. Takeo, Y. Okabe, and M. Matsuura.(2007) “Automatic seismic wave arrival detection and picking with stationary analysis: Application of the KM2O - Langevin equations”, Earth Planets Space 59, 567-577, [2007].

[11] Ali G. Hafez, T.Kohda.(2009). “Accurative P-wave Arrival Dection via MODWT”, Computer Engineering & Systems, 391 - 396 ,[2009].

[12] Pikoulis, E.-V.M. Psarakis, E.Z(2010).,”A ratio test for the accurate automatic P-wave onset detection”,Signal Processing (ICSP), 2621 – 2624, [2010]

[13] E.Choi,C.Lee,(2000)”FeatureextractionbasedontheBhattacharyya distance, Proceedings of the International Conference on Geoscience and Remote Sensing”, Hawaii, USA, 2000, pp. 2146–2148[2000].

[14] Bhattacharyya,(1943) A. "On a measure of divergence between two statistical populations defined by their probability distributions". Bulletin of the Calcutta Mathematical Society 35: 99–109. [1943].

Appendix-I

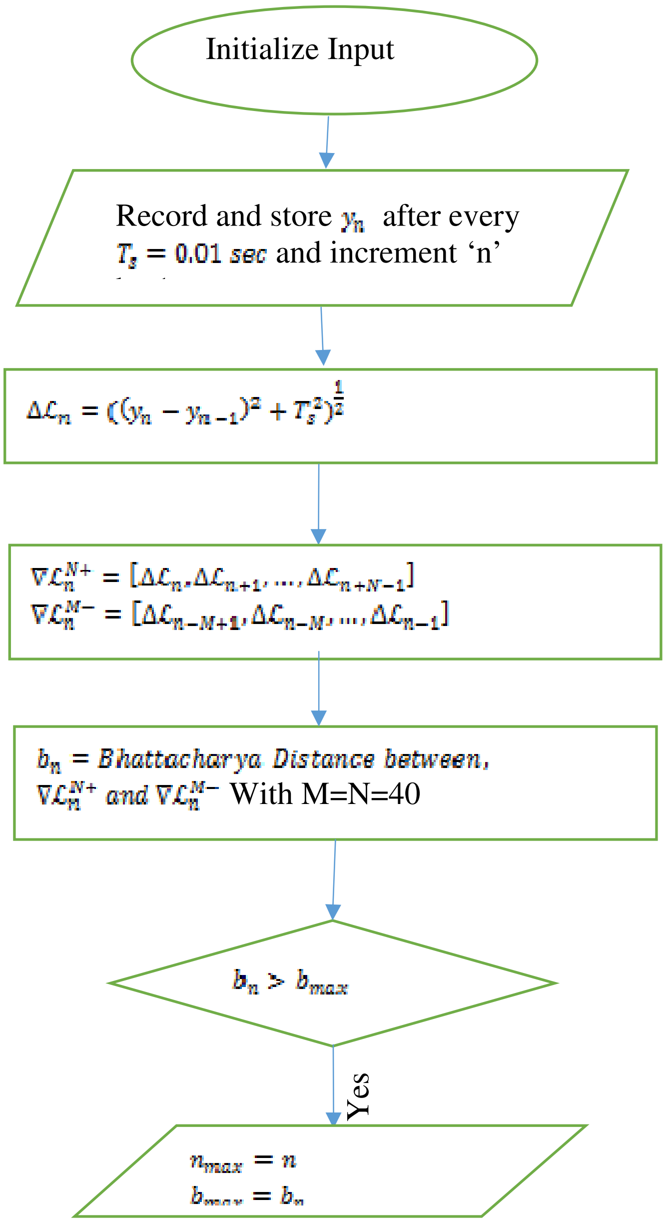

FIGURE 7: Flow chart of proposed algorithm.