International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 1217
ISSN 2229-5518

Anomaly Detection Using Hellinger Distance

Shyamal Dhua, Akash Bhosale, Nikhil Chaudhari, Saurabh Dhapre

Department of Electronics,Datta Meghe College of Enginerring, Airoli, Navi Mumbai, Maharashtra,India.
Email: shyamal.dhua@gmail.com,bhosaleakash00@gmail.com,nikhilc08@gmail.com,saurabh.2491@gmail.com

Abstract— Hellinger Distance is a distance generally used between two probability distributions.This distance is most suited to detect and measure the anisotropic nature of object parameters.Anisotropy in an object comes from the varying nature of its parameter/s in different directions.It is like, if you consider yourself to be present at the absolute centre of any object and you go on measuring the magnitude of a parameter exhibited by that object in all possible directions, then , if the object is Isotropic, on plotting a 2-d representation of your records taking the absolute centre as the origin you will most probably get a Circle. This process in the case of an Anisotropic object would have resulted in a different figure, most likely close to an Ellipse. These Anisotropic parameters residing in any object can easily be measured using Hellinger Distance. It gives us a measure of degree of anisotropy present in the object for that parameter. But, our Aim is not to measure the Hellinger Distance between Parameter variations in an object. Our Aim here is to measure the Anisotropy that results when we measure the same parameter of two objects along the same axis. The results can vary if we would have been able to implement the same process, but taking all the different directional vectors into consideration. By using this distance for Practical purposes it would become very easy to detect dissimilarity even between the most alike objects occuring in Nature. This paper discusses on the various research carried out on Hellinger distance by us along with its comparison with one of the most dominant and frequently used distance i.e. the Euclidean distance**.**

Keywords—Anisotropy; Dissimilarity; Directional vectors; Euclidean distance; Hellinger Distance; Isotropy; Parameter variations

———————————— ◆ ————————————

1 INTRODUCTION

WE propose Hellinger distance based approach to compute the divergence between two time series. With the Minimum Entropy Heuristic (MinE), the embedding dimension can automatically be chosen, thus making the algorithm a parameter-free clustering approach. For clustering time series, the similarity between two time-series is defined as the divergence between two permutation distributions.

2 WHAT IS HELLINGER DISTANCE?

The Hellinger distance is defined between vectors having only positive or zero elements. In general, it is used for row profiles. It is a type of *f*-divergence. The Hellinger distance is defined in terms of the Hellinger integral, which was introduced by Ernst Hellinger in 1909.The Hellinger distance between vectors ***P*** and ***Q*** is defined as

$$H^2(P,Q) = \frac{1}{2}\int \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda. \qquad (1)$$

d(*P*,*Q*)=[∑ (*P*^(1/2)-*Q*^(1/2))^2]^(1/2) (2)

Ernst David Hellinger

ISSN 2229-5518

Fig. 1. Ernst David Hellinger; who found Hellinger distance

3 WHAT IS ANISOTROPIC PROPERTY OF A MATERIAL?

It is the property of being directionally dependent, as opposed to isotropy which implies identical properties in all directions. It can be defined as a difference when measured along different axis.

Most other distances failed to give anisotropy between two materials to the accuracy as required as compared to Hellinger distance, so we are using Hellinger distance to find out the anisotropy between two materials. So, Hellinger distance will help us in distinguishing two materials based on their anisotropic nature. Now-a-days no such idea is present which will find out dissimilarity between two materials based on the parametric distances. So, with the help of Hellinger distance we are trying to distinguish the study material from a reference material. This project will help to detect dissimilarity between two materials where even a small difference between materials is considered to be essential.

4 HELLINGER DISTANCE BETWEEN SOUND SIGNALS

Two different sound samples were taken and the various properties of Hellinger distance and its comparison with Euclidean distance was done and the test analyses that we got were as follows:

4.1 Properties Of Hellinger Distance

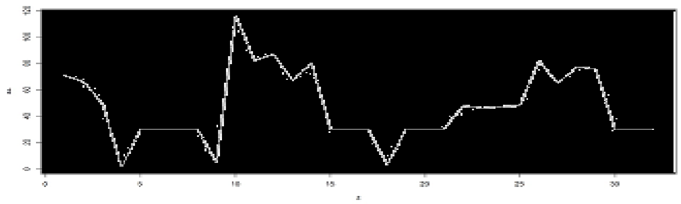
> SHIFTING PROPERTY
> TIME VARIANCE
> SENSITIVITY
> SCALING (UP AND DOWN)

4.1.1 Shifting Property:

The vectors used for this property are as follows:

>**a**
73 68 51 3 32 32 32 7 118 84 89 69 82 32 32 32 5 32 32 32 50 48 49 50 84 67 79 78 32 32 32

>**a2**
71 66 49 1 30 30 30 30 5 116 82 87 67 80 30 30 30 3 30 30 30 48 46 47 48 82 65 77 76 30 30 30



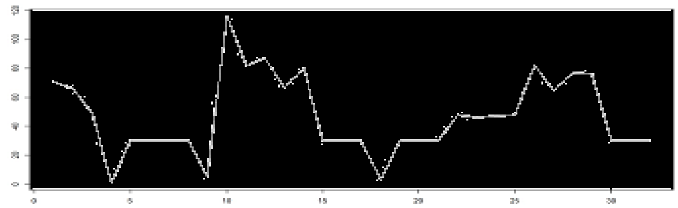


Fig. 2. Sound signals used for Shifting Property

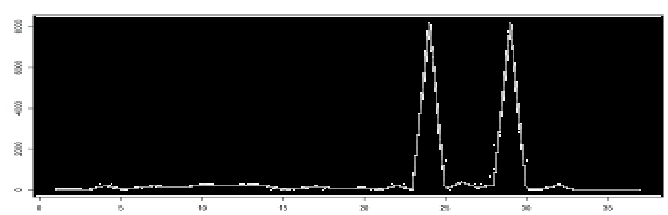
4.1.2 Time Variance:

The vectors used for this property are as follows:

> **b1**
15 26 9 203 46 89 180 127 129 235 211 198 237 196 25 66 164 80 70 89 23 214 20 8211 57 382 121 225 8225 63 60 243 0 0 0 0 0

> **b2**

0 0 0 0 0 15 26 9 203 46 89 180 127 129 235 211 198 237 196 25 66 164 80 70 89 23 214 20 8211 57 382 121 225 8225 63 60 243



ISSN 2229-5518

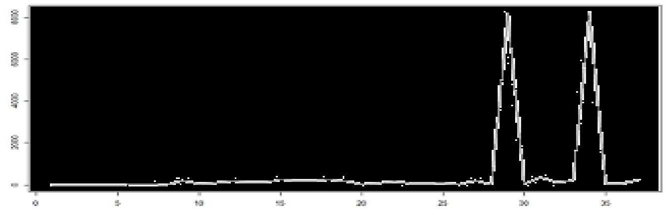


Fig. 3. Sound signals used for Time Variance Property

4.1.3 Sensitivity:

The vectors used for this property are as follows:

> **b**
15 26 9 203 46 89 180 127 129 235 211 198 237 196 25 66 164 80 70 89 23 214 20 8211 57 382 121 225 8225 63 60 243

> **c**
15 26 9 203 46 89 180 127 129 235 211 198 237 196 0 66 164 80 70 89 23 214 20 8211 57 382 121 225 8225 63 60 243

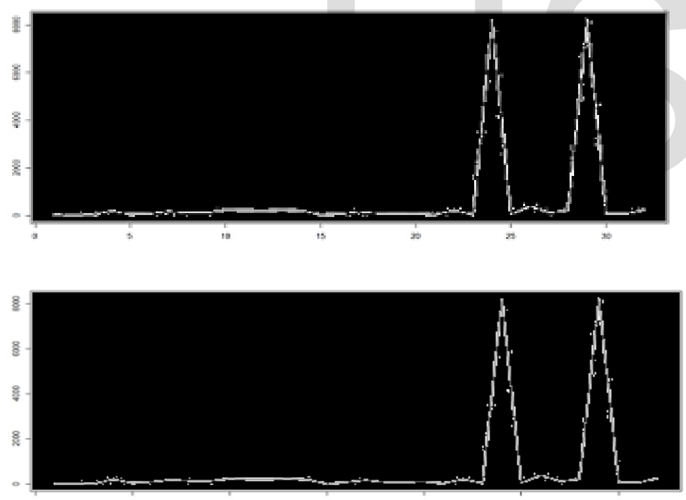


Fig. 4. Sound signals used for Sensitivity Property

4.1.4 Scaling:

The vectors used for this property are as follows:

>**Initial Signals**:

> **b2**
15 26 9 203 46 89 180 127 129 235 211 198 237 196 25 66 164 80 70 89 23 214 20 8211 57 382 121 225 8225 63 60 243

> **a2**
73 68 51 3 32 32 32 32 7 118 84 89 69 82 32 32 32 5 32 32 32 50 48 49 50 84 67 79 78 32 32 32

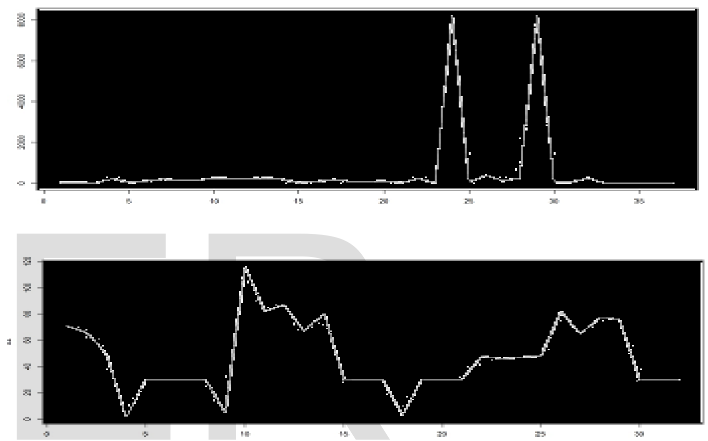


Fig. 5. Sound signals used for Scaling Property

>**Scale Up (By a Factor Of 2)**:

> **b2**
30 52 18 406 92 178 360 254 258 470 422 396474 392 50 132 328 160 140 178 46 428 40 16422 114 764 242 450 16450 126 120 486

> **a2**
146 136 102 6 64 64 64 64 14 236 168 178 138 164 64 64 64 10 64 64 64 100 96 98 100 168 134 158 156 64 64 64

>**Scale Down (By a Factor Of 0.5):**

> **b2**

7.5 13.0 4.5 101.5 23.0 44.5 90.0 63.5 64.5 117.5 105.5 99.0 118.5 98.0 12.5 33.0 82.0 40.0 35.0 44.5 11.5 107.0 10.0 4105.5 28.5 191.0 60.5 112.5 4112.5 31.5 30.0 121.5

> **a2**
36.5 34.0 25.5 1.5 16.0 16.0 16.0 16.0 3.5 59.0 42.0 44.5 34.5 41.0 16.0 16.0 16.0 2.5 16.0 16.0 16.0 25.0 24.0 24.5 25.0 42.0 33.5 39.5 39.0 16.0 16.0 16.0

4.2 Results

The results obtained from the above tests were as follows:

TABLE 1
LIST OF RESULTS OBTAINED FOR THE VARIOUS PROPERTIES

PROPERTIES	HELLINGER DISTANCE	EUCLIDEAN DISTANCE
SHIFTING	0.004973434	0.0001726361
TIME VARIANCE	0.5087497	0.1770457
SENSITIVITY	0.0195314	0.0003814755
SCALING :		
SCALE UP	0.6692311	0.3524629
SCALE DOWN	0.3346155	0.08811573

From the above table it can be clearly seen that the results obtained for Hellinger distance were better as compared to Euclidean distance.

5 AIM

The aim of our project and the corresponding hardware is to calculate a measure of the dissimilarity between two objects. The project objectives include transmission and reception of a laser beam incident on the sample and calculation of the Hellinger distance for that sample. When computed for different samples we can determine the anisotropy existing in the samples for a specific parameter.

6 PROBLEM STATEMENT

Dissimilarity detection and measurement using Hellinger Distance. The processing of the transmitted and reflected beams need to be done first. After that their respective Hellinger distance needs to be calculated and stored in memory. Using the results obtained earlier the existing anisotropy can be detected.

7 BLOCK DIAGRAM

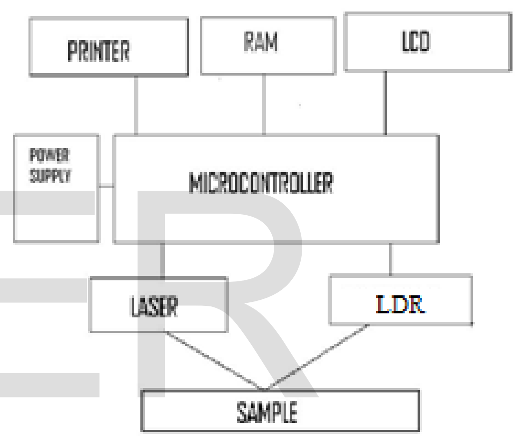


Fig. 6. Schematic of the circuit arrangement

8 SYSTEM OVERVIEW

8.1 Microcontroller

An AT89c52 Microcontroller was used due to its low cost and easy availability. Also, the 8Kbs of In-System Reprogrammable Flash Memory met our requirements. The reference sample values were stored in the Microcontroller's memory, which was used for modulation of the transmitted laser beam. Also, the calculation of Hellinger distance between the transmitted and recepted signals was done by the Microcontroller.

8.2 Laser

ISSN 2229-5518

A general purpose Red Laser diode was used for signal transmission. A DAC0808 was used for modulating the laser with the internally stored reference sample values.

8.3 LDR

As we are calculating Hellinger distance based on a specific parameter, the parameter selected by us is the absorption capacity of a sample for the laser beam. The *Triangulation* method was used to arrange the positions of the Laser and LDR for proper measurement of the reflected intensity. The reflected analog intensities are made to incident on the LDR and then using an ADC0804 we convert the recepted intensity to a digital value which is provided to the Microcontroller for processing.

8.4 Output

The output i.e. the Hellinger distance between the reference sample and the testing sample is displayed on the LCD. By using two different test samples and calculating their respective Hellinger distance and noting the results, we can easily detect the anisotropy existing between the two test samples for the same parameter.

9 CONTRIBUTION TO GENETIC STUDIES

Developing our project further can also help in integrating it with the DNA scans involved in Genetic studies. It will help in finding the dissimilarity between two long sequences of A,T,G and C present on the DNA strand. Also, by properly detecting the mutations with greater accuracy as compared to Euclidean distance, we might be able to eliminate the spread of Bio-Hazards.

10 CONTRIBUTION TO ROBOTICS

We can build a robotic dog which uses this process of dissimilarity detection. We will first provide the Reference sample's (e.g. A piece of some metal that is actually part of a bigger device) Hellinger distance as an input to the dog. The dog will then scan the entire area in search for an object whose Hellinger distance matches the one we provided it as an input, thus helping us in recovering the entire parts of our device.

11 CONCLUSION

In conclusion, the team was able to build the system and meet all of the objectives of the project in a very timely fashion. The system built by us and our research on the topic helped us in illustrating the importance and need of Hellinger distance in several important studies as compared to Euclidean distance. We can thus say that Hellinger distance is more suited as a distance for dissimilarity detection as compared to Euclidean distance.

ACKNOWLEDGMENT

First and foremost, we would like to acknowledge our supervisor, Mr. A. S. Kurhekar for putting his trust in us on this very interesting and challenging project. His encouragement, advice and motivation helped us make it through the tough times during this project. We want to also acknowledge our project guide Mr. Anil Londhe for his time, vital encouragement and support.
Last but not least, we want to thank all of our family members, friends, colleagues, and loved ones for their warm support and enthusiasm throughout the development of the project.

REFERENCES

[1]National Semiconductors Application Manual.
[2]http://paulbourke.net/miscellaneous/dft/
[3]http://www.katjaas.nl/FFTimplement/FFTimplement.html
[4]http://www.reocities.com/siliconvalley/lakes/7156/laser.html
[5]http://www.edaboard.com/thread198459.html
[6]http://forum.allaboutcircuits.com/showthread.php?t=73233
[7]http://www.rog8811.com/laserdriver.htm
[8]The figure is taken from this link http://www-history.mcs.standrews.ac.uk/history/PictDisplay/Hellinger.html