International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Metaphorical Analysis of Different Encoding Techniques for Spatial-Temporal Data

[1]Pappula Madhavi, [2]Supreethi K. P.

Submitted: 27/05/2023 **Revised**: 16/07/2023 **Accepted**: 28/07/2023

Abstract: Maps certainly plays a great way nowadays to represent spatial data. Growth of human population leads to changes in Business, environment and Society. Spatial and temporal data plays a main role to analyse and visualize current trends. Encoding is a process of keeping a given sequence of characters and converting it into a specialized and secured string format called Unicode to identify a given spatial temporal location. Different encoding techniques are used to convert data into Unicode format. This paper presents different encoding techniques performance measure with time and space complexity. It shows the results analysis of Base 32, Base 64, Elias gamma, Elias delta and Golomb codes with spatial temporal data. A comparative analysis has done among all codes and identified contingency analysis of encoding techniques. Base 64 is the the suitable and best encoding algorithm with respect to time and space on spatial and temporal data of a given point location in a trajectory data set.

***Keywords**: Spatial Temporal Data, Indexing, Base 32, Base 64, Golomb Code, Elias Gamma code, Elias Delta code*

1. Introduction

Spatial data is also known as Geo-spatial information or Geo-data, is a kind of data containing the value of geographic coordinate location of an object which exists physically, such as a building location, a road, a region and a country. We can identify an object not by its location but identified by its size,length, shape and area. Mining of Spatial-Temporal data is referred as knowledge extraction on spatial data and interesting pattern of spatial and temporal aspects. To extract spatial data, it needs coordinate point of a given location.

Mining of Spatial Temporal data is the extraction of knowledge, identifying patterns, Knowledge analysis and interesting measures in spatial and temporal exposure. Spatial temporal data mining used to find the history of lands and cities, the atmospheric conditions in weather, natural disasters like earth quakes and hurricanes and climatic changes in seasons. Spatial-temporal data mining has become more important in our life like GPS devices, map services on internet based, digital earth, satellite, Sensor and wireless networks. Many Spatial temporal mining algorithms were developed and discovered new patterns in spatial data, co-occurrence patterns and sequence of event in a particular location[1]. Many researchers has implemented and proposed many optimization techniques in spatial calculations in different scenarios. A multi layer grid model consisting of latitude and longitude is a Geo-SOT Grid model[2]. These latitude and longitude points are converting into binary form to encode into specific string format. Earth subdivision model and encoding method, first one stipulates area divided into multi-scale subdivision tiny units with equal shapes and regular sizes later method assigns unique identification string format code for each unit. Geo-spatial grid model[2,3] is widely studied in academia, effective organization and applications of large multi source spatial data.

In spatial temporal data 3D attributes like latitude, longitude and timestamp can be considered to identify the particular information of an object in a given time. All three bit steams of latitude, longitude and time are considered and convert it into 1D binary encoded form. A Space Filling Curves[5] are used to fill the adjacent space with unique code string format by using of Morton curves, Hilbert curves and Z-order curves[3]. To identify each location as unique code with encoding methods and Space filling curves are used to code and identify with neighbouring locations. Space filling curves are used to give unique codes of a given space by grid model which is divided by area subdivision methods.

Different encoding techniques are existing for data conversion from one format into another. Base 32, Base 64 encoding techniques takes required number of bits to convert it into binary format. Spatial data exists in degrees and timestamp need to be convert into epoch time. All encoding algorithms are used to convert 3D data into 1D binary format and transform it into unique code for further indexing process. Elias gamma, Elias delta and Golomb codes transform binary form into

[1]Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, Telangana - 501510, India. Email Id: p.madhavi@cvr.ac.in, ORCID ID : 0000-0002-9582-0452.

[2]Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana - 500085, India, Email Id: supreethi.pujari@jntuh.ac.in ORCID ID : 0000-0003-0110-1340

unique code generation for indexing the data. All encoding techniques transform binary bit stream into string format for indexing process.

This paper organizes section I. introduction of spatial temporal data and encoding techniques, Section II. Related work for spatial data and conversion of binary form of a given point by one example, Section III. Space filling curves representation and encoding techniques, Section IV. A comparative analysis study of all encoding techniques, Section V. Conclusion and Future work.

2. Related Work

Trajectory data is observed as the moving objects travelling history of a Person, a Vehicle, a Train or a Flight. They have both Spatial and temporal characteristics naturally contains 2D or 3D coordinates, a timestamp measure is cascading with coordinates[2,4]. A trajectory of a moving object forms a poly line in the 3D space, where two dimensions refers to spatial coordinates other dimension is Time. It can be written as series of location points with timestamped, denoted by { (p_1,q_1,t_1), (p_2,q_2,t_2), (p_3,q_3,t_3), ………(p_n,q_n,t_n) }, where p_i, q_i represents the coordinates points, t_i is the timestamp of each unit 'i' in a series of length n different locations. Geo hash is the efficient indexing method[7] for 2D location data which encode latitude and longitude information. But in ST-Hash all 3D attributes are converted into a character sequence i.e 1D ST-Hash[4] unique string.

ST- Hash Algorithm

ST-Hash algorithm[4] shows the recursive Oct tree on geographic space and it uses WGS 84 reference system for spatial coordinates, latitude ranges in between [-90^0, 90^0] and longitude scope is [-180^0, 180^0].

Figure 1 shows the encoding procedure for a given trajectory point in a given timestamp. Date and Time falls in two categories, one is leap year and non leap year. Leap year contains 527040 (366*24*60) minutes while general year contains 525600 (365*24*60) minutes. So the scope of Leap year range falls in [0 , 527040] and general year falls in [0 , 525600]. Given date and time converted into decimal and decimal to binary format is the binary bit stream of a given timestamp.

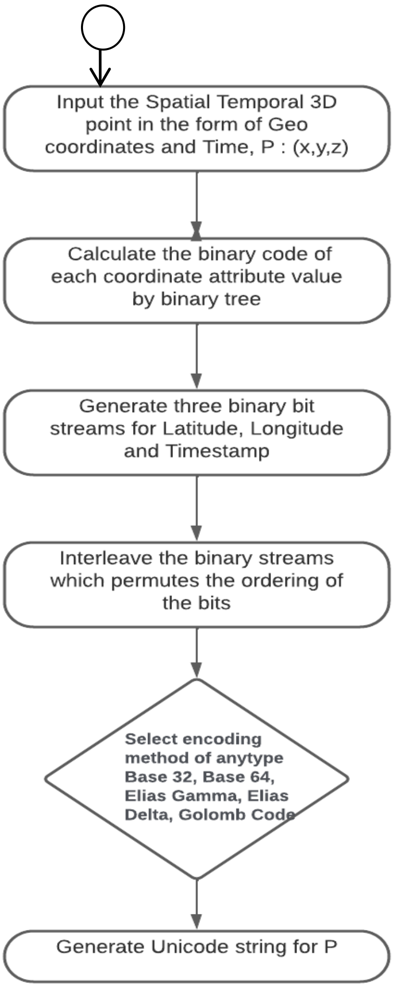

Fig 1: Flow chart for ST-Hash unique code

Latitude and Longitude[6] coordinates are converted into binary bitstream by using of binary tree construction. Total longitude range is from [-180^0 , 180^0] and the interval is split into two equal parts [-180^0, 0^0] [0^0 , 180^0], it is a first division. If a given coordinate is falls into left block, write bit 0 in bitstream object and if coordinate falls into right partition then write bit 1 into bit stream for first division. Later do equal partition of all intervals in second division then write the 1s or 0s depending on the coordinate falling interval. Divide the binary tree into third, fourth, fifth division….and so on continue the process when the given coordinate point falls into a small range interval. Cascade all bits one by one by each division.

Figure 2: shows the binary tree for writing binary stream for a given longitude location. Latitude[6] falls into [-90^0 , 90^0] range, apply same procedure to divide it into two parts in first division as [-90^0 , 0^0] [0^0 , 90^0]. Later all divisions divide each interval into two equal parts, write the binary stream of a given latitude point. ST-Hash algorithm is used to represent latitude and longitude points into binary stream.

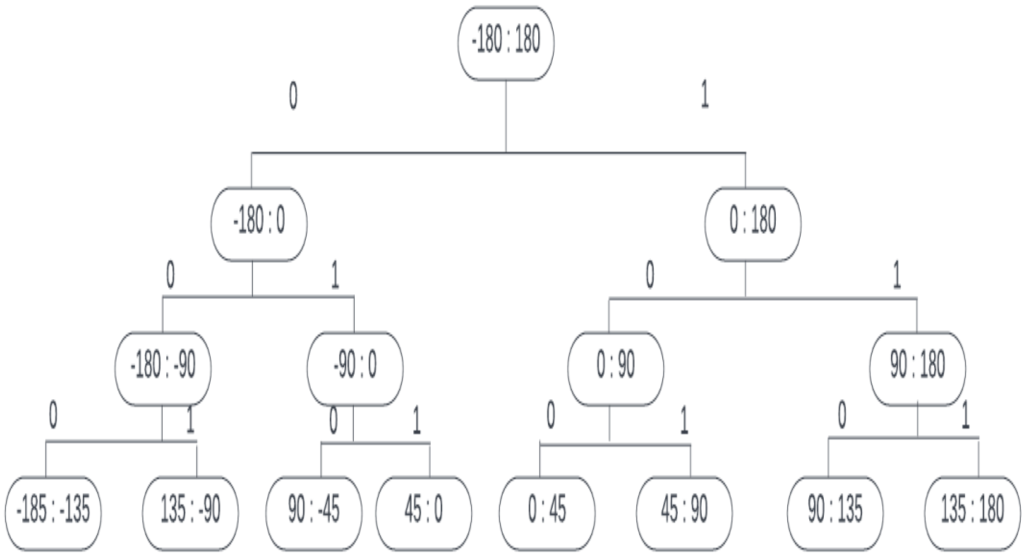

Fig 2: Binary tree construction for a given coordinate point of Longitude.

For example binary form of a given point of Spatial and temporal (17^0 , 78^0, 2022-March-10, 09:30:25 AM) is as follows

Latitude : 1000110000

Longitude : 1011011110

Time : 0100110001

Time is converted into epoch timestamp[6] in seconds i.e 1646904625 and then converted into binary form. Now three bit streams are interleaved and write it into 1D bit stream.

1D stream : 110001010010101111010010010001

Base 32 string conversion :

Binary code : 11000 10100 10101 11101 00100 10001

Decimal Value : 24 20 21 29 4 17

Char String : Y U V 5 E R

Unicode for the given location of a particular time is : “YUV5ER”.

Base 64 string conversion :

Binary code : <u>110001</u> <u>010010</u> <u>101111</u> <u>010010</u> <u>010001</u>

Decimal Value : 49 18 47 18 17

Char String : x S v S R

Unicode for the given location of a particular time is : “xSvSR”.

Elias Gamma code conversion :

Binary code : 110001 010010 101111 010010 010001

Decimal Gamma code : 9 18 31 18 17

Hex Code : 9 12 1F 12 11

Unicode for the given location of a particular time is : "9 12 1F 12 11".

Elias Delta code conversion :

Binary code : 110001 010010 101111 010010 010001

Decimal Delta code : 35 37 31 5 3

Hex Code : 23 25 1F 5 3

Unicode for the given location of a particular time is : " 23 25 1F 5 3 ".

Golomb code conversion :

Binary code : 110001 010010 101111 010010 010001

Decimal Delta code : 16 17 14 17 16

Hex Code : 10 11 E 11 10

Unicode for the given location of a particular time is : "10 11 E 11 10".

3. Space Filling Curves (SFC)

The generated unique code is a key to identify the desired location in a particular time. All keys are indexed in a tree into hierarchical structure which is able to perform the fast disk access. To give Unicode per each location and keep all codes in ordered sequence, we need to use Space Filling Curves (SFC). SFCs are used to identify the adjacent points for analysis and comparison in real time applications. SFC is a paradigm mostly used for ordering multidimensional data[5,11]. SFCs are used in many applications for set of points in an ordered manner.

Recent days COVID-19[3,12] analysis for adjacent cities with respect to number of COVID effected people, number of recovered people from COVID, Number of deaths occurred in each city can be compared by Space Filling Curves (SFC) easily.

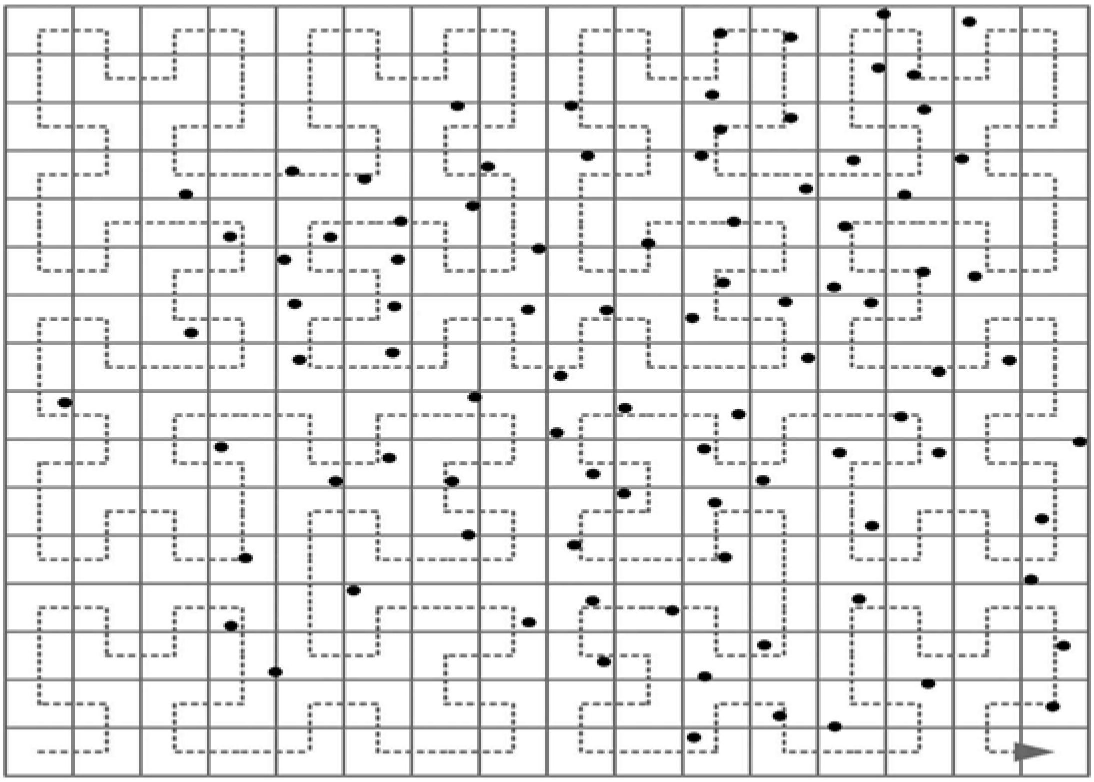

Fig 3: Hilbert curve 2D representation of level 3 (16*16)

For example 2D space partition into a subdivision grid, such that each point lies in a cell. Hilbert curve[5] plays a main role to connect cells in H passion. 3D Hilbert curves are represented with three coordinate points P(X, Y, Z) to analyze the data in detail passion with a given order m. Encoding and decoding can be done in a intuitive iteration[6]. Based on locations of object state encoding and decoding algorithms are proposed for generating codes. SFCs are used to generate a code for neighbouring points in a sequential order to index the data.

Base 32 Encoding scheme has taken for each Five (2^5) bits from LSB(Least Significant Bit) of a given 1D bit stream. Padding can be done with required number of bits for MSB (Most Significant Bit). Generate Unicode of a given location point from Base 32 table corresponding to decimal value[4,8]. Base 32 table is as follows for a given decimal value and write string code corresponding to a given code.

Base 64 code has taken Six (2^6) bits and generated a Unicode string[9]. Hex code is used for Elias gamma, Delta and Golomb Coding techniques. All five different algorithms are applied to a data set of Porto Taxi data set partial trajectories [9] of a given Taxi Trip ID. A comparative analysis has done with all five algorithms with space and time complexity. Among all encoding techniques Base 64 has takes less time and less space to index spatial and temporal data.

Elias Gamma encoding is an universal code for positive integers for which upper bound can not be determined before. It is used in many applications where not knowing about largest encoded value or used to compress data for small values which are more frequent[11]. Time and Space complexity is higher than Base 32 and Base 64 encoding algorithms.

Elias Delta Encoding[11] scheme is an universal code for positive integers by peter Elias. It does not work for Negative numbers and Zero. Elias Delta occupies more storage and to generate Unicode it takes more time to compare with all encoding algorithms. Elias Delta encoding algorithm takes highest time and storage comparing to all algorithms.

Golomb Code is a lossless data compression method to encode the data of a given location point. It uses a tuned parameter to divide an input value x into two parts q, the results of divisor M and r, the remainder.

All encoding algorithms generate a transformation code of a given point. Data must be encoded into a specified format and decoded to get the original data. Encoding and decoding techniques used to convert data into another form and get the original data back.

4. Comparative Study

Spatial and temporal data of Porto Taxi data set for partial trajectories[10] has been taken from UCI Machine learning repository and applied data set for all different algorithms of Base 32, Base 64, Elias Gamma, Elias Delta and Golomd code. Dataset contains 327 taxi ids for different trips to get the trajectory path. Different locations are identified and converted into Unicode by using encoding schemes for each taxi trip. Taxi trips has taken for 327 rows with different taxi ids. Each trip_id identified a trajectory path of each location and generated a uni code of spatial and temporal data. Each trip id generate Unique code for every 15 seconds. Time and Space complexities has monitored for all encoding schemes. Average time and space has taken for comparative analysis

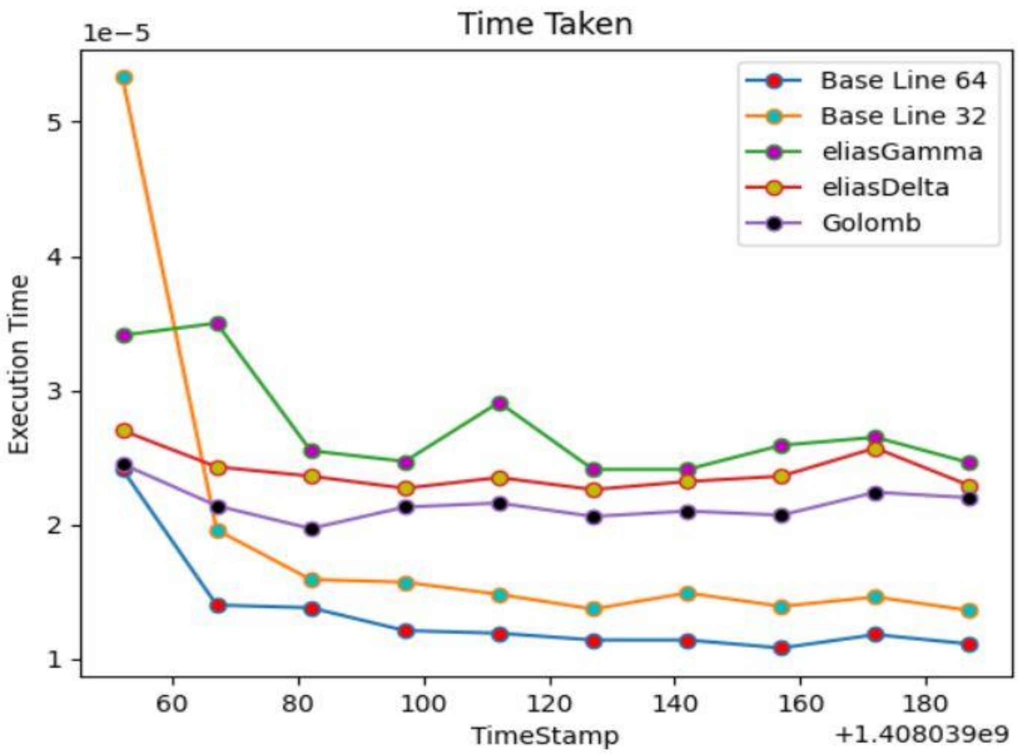

Fig 5: Time complexity of different encoding algorithms for Trip_id

Time complexity of all encoding methods shown in figure 5. Each line graph indicates its trajectory path of different locations. Execution time to convert location point into unique code is high in Elias gamma method. Average execution time for Base 64 is minimum and can be considered for spatial indexing. My research analysis can be preferred to consider Base 64 encoding method to convert spatial temporal location into binary stream. Later indexing can be done with large data and accessing time is less to know about any particular point where the taxi is. This comparison analysis identified the best encoding scheme with respect to execution time and storage space. Average time complexity is taken to execute each trip id and used for further analysis.

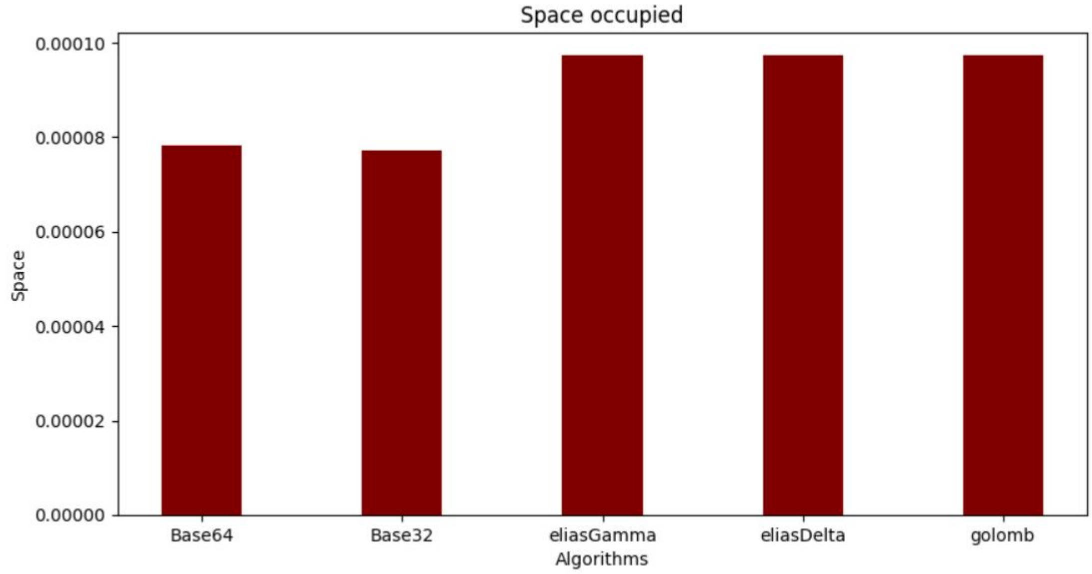

Fig 6: Space complexity of different encoding algorithms for Trip_id

Storage size of each trip has taken and visualized on Bar chart. Among all encoding schemes Base 32 and Base 64 has equal size comparing to Elias gamma, Elias Delta and Golomb Codes. Storage space required for elias gamma is more comparing with other encoding techniques. Average data stored for each trip id is more while Base 62 or Base 64 can be taken as it requires less storage.

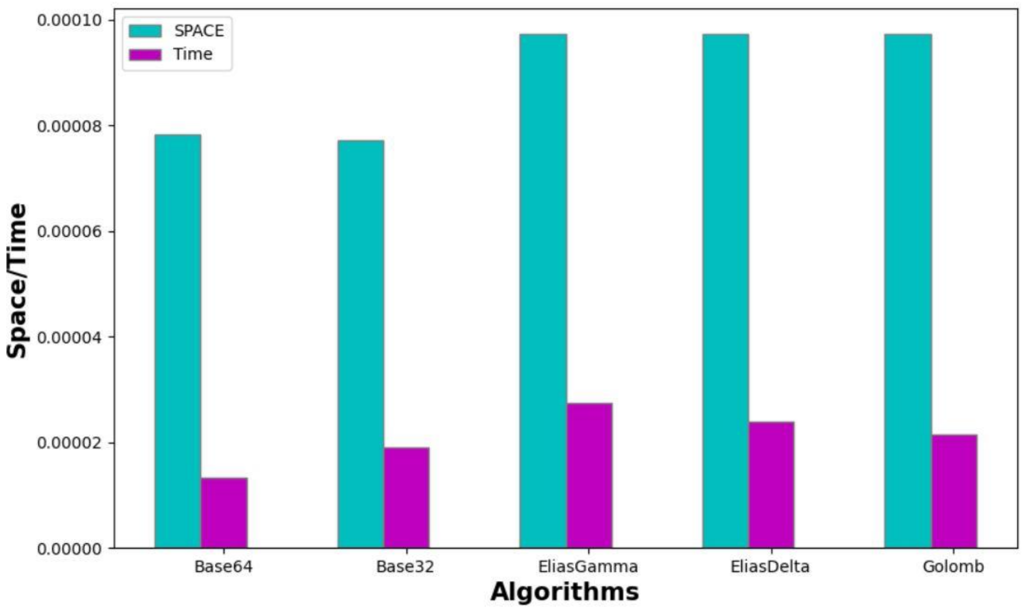

Fig 7: Average Space and Time complexity of a Taxi for Trip_id

Average time and storage has taken for all encoding schemes for each trip id. The above bar graph shows the Space complexity and Time complexity for different algorithms. This work concluded that to generate 1D data for 3D spatial and temporal data, we may use Base 64 encoding scheme for further works because it has less Space and Time complexity comparing with remaining encoding techniques.

Indexing of all unique codes generated by different taxi ids gives fast disk access if accessing of any required data. Unique codes are used for indexing for spatial and temporal data.

5. Conclusion and Future work

Spatial and temporal data generates a large voluminous data size Recent Days. To store and access large data we need indexing in proper manner. R trees are used to index spatial and temporal data in high accuracy. To index large spatial data, we need to generate a Unicode to compress large bit stream. This paper focused on all encoding techniques for large spatial data to generate 1D bit stream for 3D data. Unique codes are generated and need to be indexed for fast access. A comparative analysis has done for all Encoding techniques and identified a best technique with less time and space complexity. In future a new method may be implemented to compress all unique codes which requires a very less storage and execution time. To handle and generate unique codes for large bit stream, new compression methods may require.

References

[1] Aydin, Berkay & Angryk, Rafal. (2018). A Gentle Introduction to Spatiotemporal Data Mining. 10.1007/978-3-319-99873-2_1.

[2] Hou, K.; Cheng, C.; Chen, B.; Zhang, C.; He, L.; Meng, L.; Li, S. A Set of Integral Grid-Coding Algebraic Operations Based on GeoSOT-3D. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 489. https://doi.org/10.3390/ijgi10070489.

[3] Madhavi Pappula, and K. P. Supreethi. "A QUALITATIVE STUDY OF SPATIO-TEMPORAL INDEXING TECHNIQUES FOR GEO-SPATIAL DATA: A REVIEW ON SPATIO-TEMPORAL INDEXING METHODS." Journal of Tianjin University Science and Technology ISSN (Online): 0493-2137 E-Publication: Online Open Access Vol:54 Issue:07:2021 DOI 10.17605/OSF.IO/ C4M3J.

[4] X. Guan, C. Bo, Z. Li and Y. Yu, "ST-hash: An efficient spatiotemporal index for massive trajectory data in a NoSQL database," 2017, 25th International Conference on Geoinformatics, 2017, pp. 1-7, doi: 10.1109/GEOINFORMATICS.2017.8090927.

[5] Narimani Rad, Hossein & Karimipour, Farid. (2019). Representation and generation of space-filling curves: a higher-order functional approach. Journal of Spatial Science. 66.1-21. 10.1080/14498596.2019.1668870.

[6] Ježek, J., Kolingerová, I. (2014). STCode: The Text Encoding Algorithm for Latitude/Longitude/Time. In: Huerta, J., Schade, S., Granell, C. (eds) Connecting a Digital Europe Through Location and Place. Lecture Notes in Geoinformation and Cartography. Springer, Cham. https://doi.org/10.1007/978-3-319-03611-3_10

[7] JIA, L., LIANG, B., LI, M., LIU, Y., CHEN, Y. and DING, J. (2022), Efficient 3D Hilbert Curve Encoding and Decoding Algorithms. Chinese Journal of Electronics, 31: 277-284. https://doi.org/10.1049/cje.2020.00.171.

[8] Zhang, Zekun & Sun, Xiaoting & Chen, Siyang & Liang, Yongquan. (2022). LPPS-AGC: Location Privacy Protection Strategy Based on Alt-Geohash Coding in Location-Based Services. Wireless Communications and Mobile Computing. 2022. 1-17. 10.1155/2022/3984099.

[9] Shaamood, Mohammed Thakir.(2021), "Encoding JSON by using Base 64", Iraqi Journal for Electrical & Electronic Engineering . Jun2021, Vol. 17 Issue 1, p29-37,DOI: 10.37917/ijeee.17.1.4.

[10] Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., "Predicting Taxi–Passenger Demand Using Streaming Data". In: IEEE Transactions on Intelligent Transportation Systems, vol.14, no.3, pp.1393-1402, September (2013).

[11] Elias, Peter (March 1975). "Universal codeword sets and representations of the integers". IEEE Transactions on Information Theory. 21 (2): 194–203. doi:10.1109/tit.1975.1055349.

[12] Ali ME, Eusuf SS, Islam KA. An Efficient Index for Contact Tracing Query in a Large Spatio-Temporal Database. arXiv preprint arXiv:2006.12812. 2020 Jun 23.

[13] Thota, M. K. ., P M, P. ., & K R, V. . (2023). Proffering Ranks to the Smart Cities based on the Data Received from IoT Devices using Visualization Techniques. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 204–214. https://doi.org/10.17762/ijritcc.v11i4.6402

[14] Mwangi, J., Cohen, D., Silva, C., Min-ji, K., & Suzuki, H. Feature Extraction Techniques for Natural Language Processing Tasks. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/137

[15] Agrawal, S. A., Umbarkar, A. M., Sherie, N. P., Dharme, A. M., & Dhabliya, D. (2021). Statistical study of mechanical properties for corn fiber with reinforced of polypropylene fiber matrix composite. Materials Today: Proceedings, doi:10.1016/j.matpr.2020.12.1072