

Similarity Pattern Search for Stream Time Series Image Data using Cluster Median Approach

Vishwanath R Hulipalled^a, K Shreekrishna Kumar^b, Venugopal K R^a, S S Iyengar^c and L M Patnaik^d

^aDepartment of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. Contact: vishwa.gld@gmail.com

^bDirector, All India Council for Technical Education, SWRO, Bangalore, India

^cDirector and Ryder Professor, Florida International University, USA

^dHonorary Professor, Indian Institute of Science, Bangalore 560 001, India

Stream Time Series retrieval has been a major area of study due to its vast application in various fields like weather forecasting, multimedia data retrieval and huge data analysis. Its original task is to identify those time series similar to pattern (query) time-series image data, where both pattern and time series image data are static. Presently, there is a demand for image data stream processing, quick searching and fast response of such online data. In this paper, we use a cluster median or segment median method for similarity matching between static/dynamic patterns and stream time series image data which can be computed with respect to change in temporal events and hence suitable for the temporal behaviour. Mainly, we present an effective pruning technique on the multilevel representation of the image data. The objective of pruning technique is to reduce the search space and to retrieve the similar patterns very efficiently. Experiments show that our approach *MCM* performs well compared to existing methods *DWT* and *MSM*.

Keywords: Cluster Median, Multilevel Segment, Similar Pattern match, Stream Time Series

1. INTRODUCTION

Time series is a sequence of values or events obtained over repeated measurements of time. The values are typically measured at equal intervals of time *e.g.*, monthly, quarterly, half yearly, yearly, *etc.*, In traditional time series data processing, regular or frequent updates of the data is the most difficult task. Hence there is a necessity of new approach in the management of Stream Time-Series data.

Stream time-series is continuously generated massive stream of data which is ordered and rigourously changing. Stream time-series is used in large spectrum of applications, such as surveillance of sensor network[1], industry production process, electric power grid, moving object search[2][3], financial data analysis[4], internet traffic analysis[5] *etc.* In the dynamic

environments where there is tremendous and potentially infinite volumes of data streams, there is a need for incremental methods that are able to update changing models and strategies to identify and manage patterns of time series.

1.1. Image Stream Time-Series

The primary types of objects in our stream processing framework are images and streams. In Image Algebra [6], an image consists of two things, a set of points in some n -dimensional space and a set of values associated with points.

Definition – 1(PointSet)- A point set is some topological space, consisting of points and a topology that provides for notions such as distance between two points and neighborhood of a point.

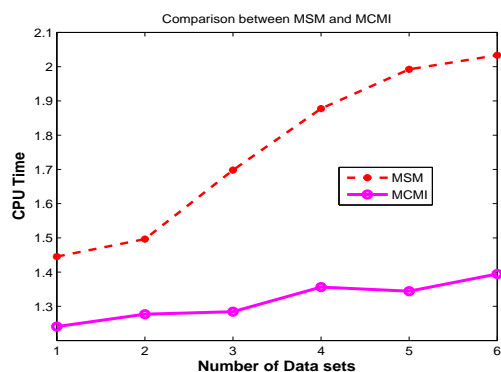


Figure 5. Comparison of MSM and MCFI Algorithms

7. CONCLUSIONS

In this paper, we propose a multilevel cluster median of stream time series image data representation method for pattern matching. The proposed *MCFI* representation can be adopted for the temporal nature of the data stream to achieve the fast similarity search. High speed pattern match is performed on stream time series using the *MCFI* Multilevel Cluster Median Image data approximation, where the data is pruned and computational cost is reduced. The results obtained by our algorithm Multilevel Cluster Median Image *MCFI* is better than the existing algorithms such as Discrete Wavelet Transform (*DWT*) and Multiscale Segment Mean (*MSM*) [12]. This approach can be further applied to multi-dimensional data like streaming video data and multimedia data.

REFERENCES

1. M Li and Y Liu. Underground Coal Mine Monitoring with Wireless Sensor Networks, *ACM Transactions Sensor Networks*, 2009.
2. L Chen, M T Ozsu, and V. Oria, Robust and Fast Similarity Search for Moving Object Trajectories, *Proceedings of ACM SIGMOD*, Pages 491-502, 2005.
3. Y Liu, L Xiao, X Liu, L M Ni and X Zhang. Location Awareness in Unstructured Peer-to-Peer Systems, *IEEE Transactions Parallel and Distributed Systems*, 16(2):163-174, Feb. 2005.
4. H Wu, B Salzberg, and D Zhang. Online Event Driven Subsequence Matching over Financial Data Streams, *Proceedings of ACM SIGMOD*, Pages 23-29, 2004.
5. C Cranor, T Johnson, and O Spatscheck. Gigascope: A Stream Database for Network Applications, *Proceedings of ACM SIGMOD*, Pages 647-651, 2003.
6. G X Ritter, J N Wilson. Handbook of Computer Vision Algorithms in Image Algebra (2nd Edition), *CRC Press*, Pages 1-417, 2001.
7. A Kemper and B Stegmaier. Evaluating Best-matchjoins on Streaming Data, *Technical Report MIP-0204, Universitt Passau, 2002. Algorithms (FODO)*, 1993.
8. R Agrawal, K I Lin, H S Sawhney, and K Shim. Fast Similarity Search in the presence of Noise, Scaling, and Translation in Time-Series Database, *Proceedings of 21st International Conference. Very Large Data Bases (VLDB)*, Pages 490-501, 1995.
9. R Agrawal, C Faloutsos and A N Swami. Efficient Similarity Search in Sequence Databases, *Proceedings of 4th International Conference. Foundations of Data Organization and Algorithms (FODO)*, Pages 69-84, 1993.
10. C Faloutsos, M Ranganathan and Y Manolopoulos. Fast Subsequence Matching in Time-Series Databases, *Proceedings of ACM SIGMOD*, Pages 419-429, 1994.
11. A Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching, *Proceedings of ACM SIGMOD*, Pages 47-57, 1984.
12. Xiang Lian, Lei Chen, Jeffrey Xu Yu, Jinsong Han, Jian Ma. Multiscale Representations for Fast Pattern Matching in Stream Time Series, *IEEE Transaction on Knowledge and Data Engineering*, 21(4):568-581, April 2009.
13. R Sethukkarasi, D Rajalakshmi and A Kannan. Efficient and Fast Pattern Matching in Stream Time Series Image Data, *Proceedings of First International Conference on Integrated Intelligent Computing*, Pages 130-135, 2010.
14. L Chen and R Ng. On the Marriage of Edit Distance and Lp Norms, *Proceedings of 30th International Conference on Very Large Data Bases (VLDB)*, Pages 227-234, 2004.
15. B K Yi, H Jagadish and C Faloutsos. Efficient Retrieval of Similar Time Sequences under Time Warping, *Proceedings of 14th International Conference on Data Engineering (ICDE)*, Pages 201-208, 1998.

16. E Keogh. Exact Indexing of Dynamic Time Warping, *Proceedings of 28th International Conference on Very Large Data Bases (VLDB)*, Pages 406-417, 2002.
17. J S Boreczky and L A Rowe. Comparison of Video Shot Boundary Detection Techniques, *Proceedings of 8th International Symposium. Storage and Retrieval for Image and Video Databases*, Pages 170-179, 1996.
18. M Vlachos, G Kollios and D Gunopulos. Discovering Similar Multidimensional Trajectories, *Proceedings of 18th International Conference on Data Engineering (ICDE)*, 2002.
19. V Megalooikonomo, Q Wang, G Li and C Faloutsos. A Multiresolution Symbolic Representation of Time Series, *Proceedings of 21st International Conference on Data Engineering (ICDE)*, Pages 668-679, 2005.
20. Y Zhu and D Shasha. Warping Indexes with Envelope Transforms for Query by Humming, *Proceedings ACM SIGMOD*, Pages 181-192, 2003.
21. Y Zhu and D Shasha. Efficient Elastic Burst Detection in Data Streams, *Proceedings of 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Pages 336-345, 2003.
22. A Bulut and A K Singh. A Unified Framework for Monitoring Data Streams in Real Time, *Proceedings of 21st International Conference on Data Engineering (ICDE)*, Pages 44-55, 2005.
23. L Gao and X S Wang. Continually Evaluating Similarity-Based Pattern Queries on a Streaming Time Series, *Proceedings ACM SIGMOD*, Pages 370-381, 2002.
24. S Papadimitriou, J Sun and C Faloutsos. Streaming Pattern Discovery in Multiple Time-Series, *Proceedings of 31st International Conference on Very Large Data Bases (VLDB)*, Pages 697-708, 2005.
25. B K Yi, N Sidiropoulos, T Johnson, H V Jagadish, C Faloutsos and A Biliris. Online Data Mining for Co-Evolving Time Sequences, *Proceedings of 16th International Conference on Data Engineering (ICDE)*, Pages 13-22, 2000.
26. H Wu, B Salzberg and D Zhang. Online Event Driven Subsequence Matching over Financial Data Streams, *Proceedings ACM SIGMOD*, Pages 322-331, 2004.



Vishwanath R Hulipalled

is an Assistant Professor in the Department of Computer Science and Engineering at Sambhram Institute of Technology, Bangalore, India. He received his Bachelors degree in Computer Science and Engineer-

ing from the Karnataka University and Master of Engineering from UVCE, Bangalore University, Bangalore. He is presently pursuing his Ph.D in the area of Data Mining in JNTU Hyderabad. His research interest includes Time Series Mining and Financial Data Analysis.



K. Shreekrishna Kumar

is currently the Director of All India Council for Technical Education, SWRO, Bangalore. He obtained his Master of Science from Bhopal University. He received his Masters degree in Information Technology from Punjabi University. He was awarded

Ph.D. in Physics (Glass Technology) from Mahatma Gandhi University. He was the member of the Jury Panel, Indian Journal of Pure and Applied Physics, CSIR (New Delhi), He was the Collaborative researcher, Nuclear Science Centre, New Delhi.

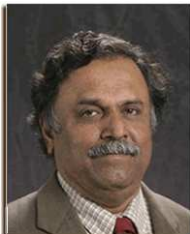


Venugopal K R

is currently the Principal, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. He received his Masters degree in Computer Science and

Automation from Indian Institute of Science Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored and edited 35 books on Computer Science and Economics, which include Petrodollar and the World Economy, C

Aptitude, Mastering C, Microprocessor Programming, Mastering C++ and Digital Circuits and Systems etc.. During his three decades of service at UVCE he has over 300 research papers to his credit. His research interests include Computer Networks, Wireless Sensor Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.



S S Iyengar is currently the Roy Paul Daniels Professor and Chairman of the Computer Science Department at Louisiana State University. He heads the Wireless Sensor Networks Laboratory and the Robotics Research Laboratory at LSU.

He has been involved with research in High Performance Algorithms, Data Structures, Sensor Fusion and Intelligent Systems, since receiving his Ph.D degree in 1974 from MSU, USA. He is Fellow of IEEE and ACM. He has directed over 40 Ph.D students and 100 Post Graduate students, many of whom are faculty at Major Universities worldwide or Scientists or Engineers at National Labs/Industries around the world. He has published more than 500 research papers and has authored/co-authored 6 books and edited 7 books. His books are published by John Wiley & Sons,

CRC Press, Prentice Hall, Springer Verlag, IEEE Computer Society Press etc.. One of his books titled Introduction to Parallel Algorithms has been translated to Chinese.



L M Patnaik is currently Honorary Professor, Indian Institute of Science, Bangalore, India. He was a Vice Chancellor, Defense Institute of Advanced Technology, Pune, India and was a Professor since 1986 with the Department of Computer Science and Automation, Indian

Institute of Science, Bangalore. During the past 35 years of his service at the Institute he has over 700 research publications in refereed International Journals and refereed International Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. His areas of research interest have been Parallel and Distributed Computing, Mobile Computing, CAD for VLSI circuits, Soft Computing and Computational Neuroscience.