Mining and Linking Patterns Across Live Data Streams and Stream Archives

Di Yang, Kaiyu Zhao, Maryam Hasan, Hanyuan Lu, Elke A. Rundensteiner, Matthew O. Ward
Computer Science Department - xmdv@cs.wpi.edu

Motivation

Streaming Data Everywhere Carries Hidden Insights

Highway Traffic   Online Transactions   Network Monitoring

Data Streams

Need for interactive exploration of complex patterns

System Objectives

Pattern Mining: Highly efficient mining of complex patterns within live data streams and stream pattern archives.
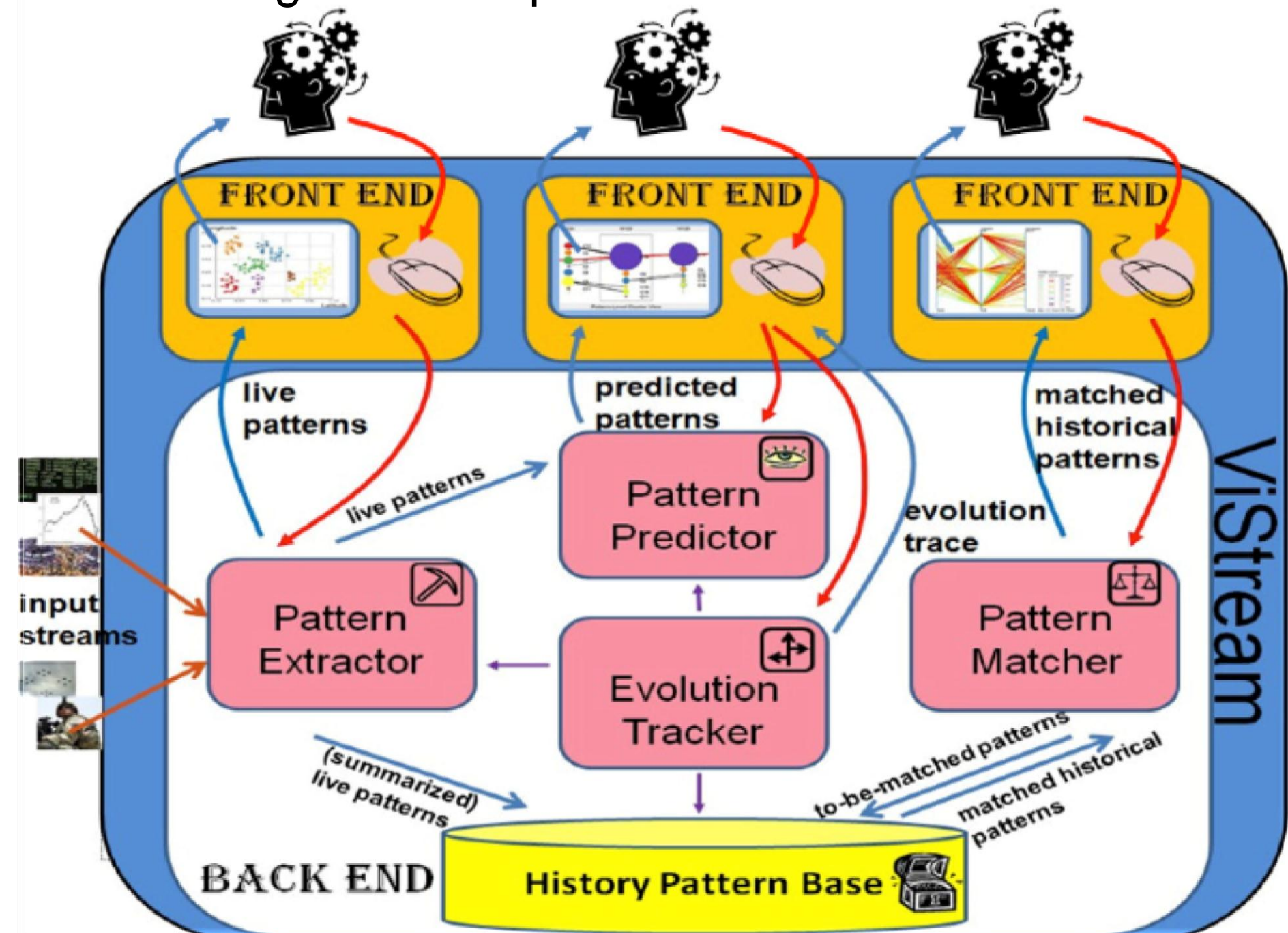
Summarization: Summarize extracted patterns into descriptive yet highly compact formats by employing multi-resolution compression strategies.

Evolution tracking: Track evolution of pattern changes over time from past, to present and even to the near future.

Visualization: Intuitively display the mining results to human analysts through visual displays and interactions.

ViStream System Architecture

Rich integrated computational and visual functionalities

Pattern Extraction and Summarization

Pattern Extraction: Developed innovative pattern detection algorithms [1,2,5] to efficiently extract interesting patterns from data streams.
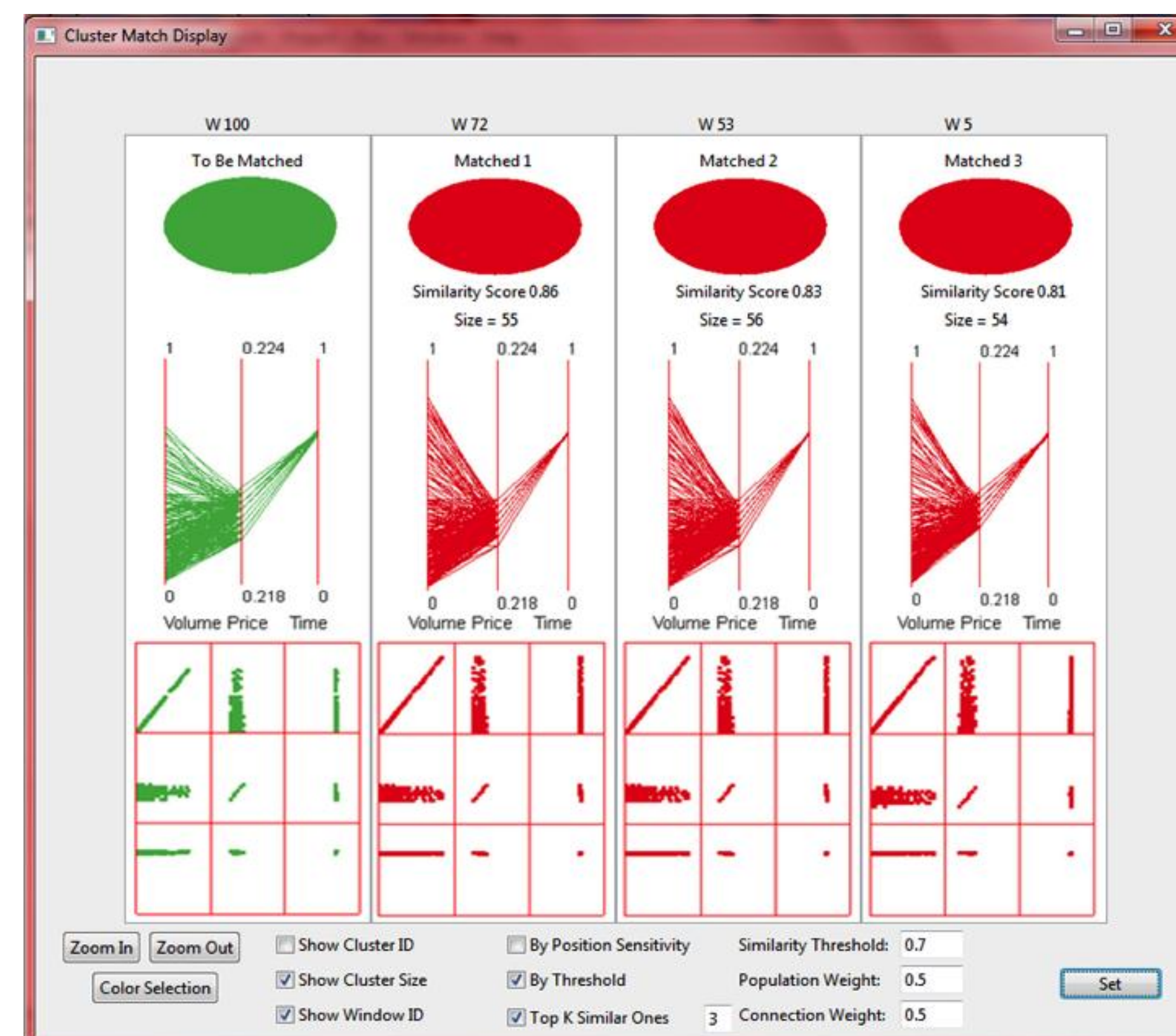
Pattern Summarization: Designed novel multi-resolution compression strategies that compress these extracted patterns into descriptive yet highly compact formats at stream speed [4]

Raw Cluster (Full Representation)   Basic SGS   Compressed SGS
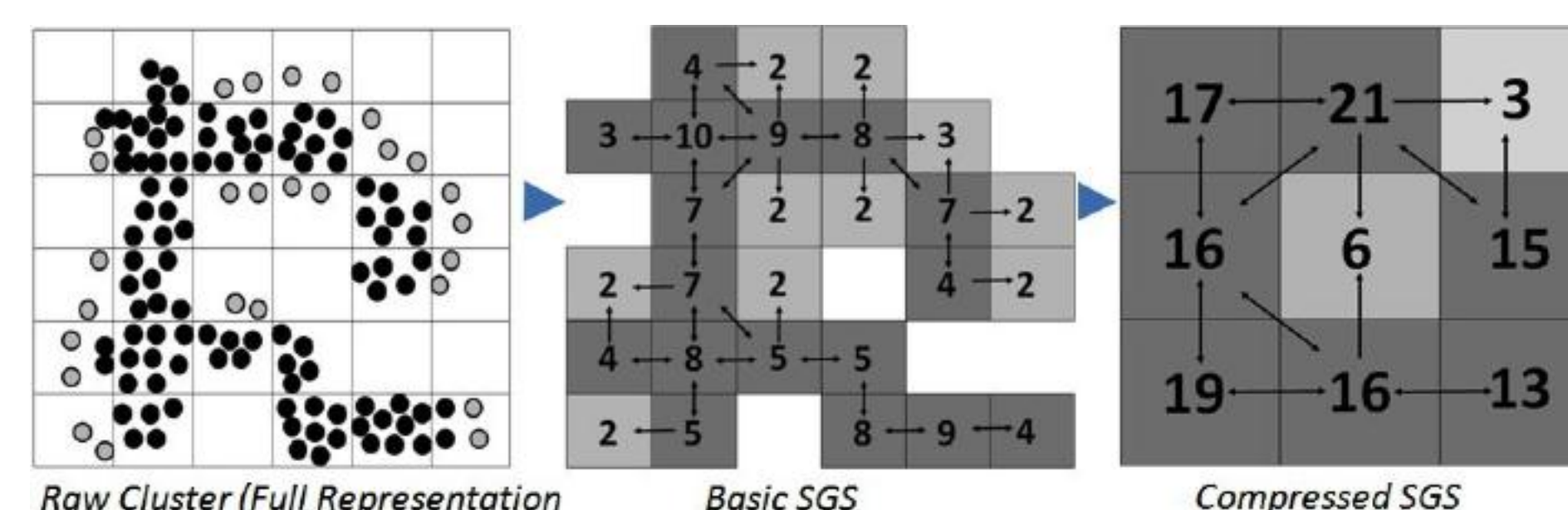
Cluster compression example at 2 levels

Pattern Matching

Pattern Matching: Developed matching algorithms that, given a pattern of interest in the current stream window, identify matching ones from historical pattern archive at any resolution [4]
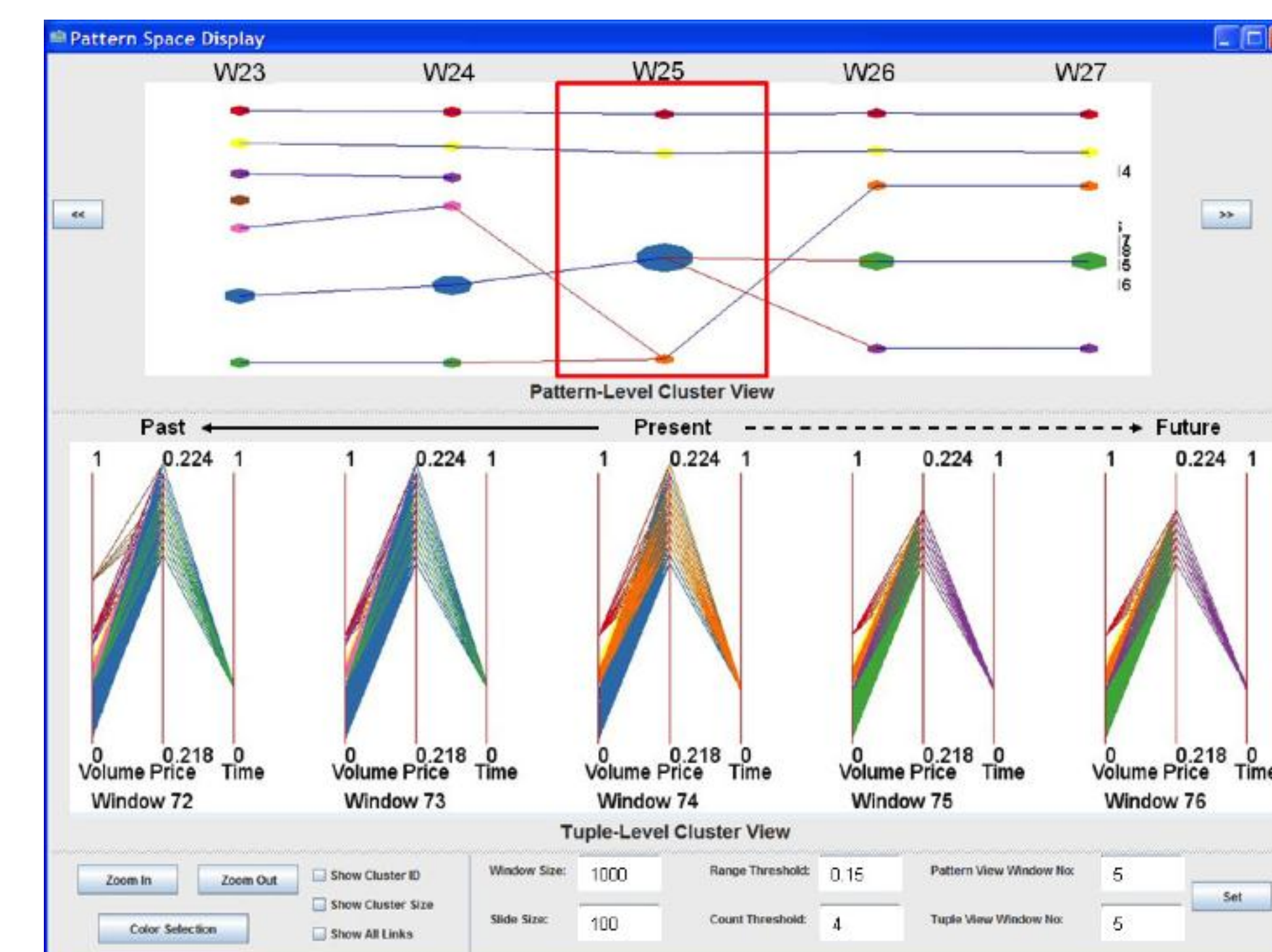

Stock price data

Evolution Tracking

Evolution Tracking: monitors interrelationships between historical, live and prospective patterns to determine evolution of patterns over time and other trends [3]

Stock price data

Results and Contributions

- ViStream tool supports interactive pattern exploration for mining, compressing & querying stream patterns
- Experimental studies demonstrate scalability of pattern extraction and summarization techniques
- User studies confirm the effectiveness of technology for high-quality compression and matching
- Release of ViStream technology as freeware software

References

[1]. Di Yang, Elke A. Rundensteiner, Matthew O. Ward: Mining neighbor-based patterns in data streams. Inf. Syst. 38(3): 331-350 (2013)

[2]. Di Yang, Elke A. Rundensteiner, Matthew O. Ward: Shared execution strategy for neighbor-based pattern mining requests over streaming windows. ACM Trans. Database Syst. 37(1): 5 (2012)

[3]. Yang, Di, Zhenyu Guo, Elke A. Rundensteiner, and Matthew O. Ward. "CLUES: a unified framework supporting interactive exploration of density-based clusters in streams." In *Proceedings of the 20th* ACM CIKM, 815-824. (2011)

[4]. Di Yang, Elke A. Rundensteiner, Matthew O. Ward: Summarization and Matching of Density-Based Clusters in Streaming Environments. PVLDB 5(2): 121-132 (2011)

[5]. Di Yang, Elke A. Rundensteiner, Matthew O. Ward: A Shared Execution Strategy for Multiple Pattern Mining Requests over Streaming Data. PVLDB 2(1): 874-885 (2009)

This is a demo poster for VLDB 2013

Software released @ davis.wpi.edu/xmdv