HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

SIATEC-C: COMPUTATIONALLY EFFICIENT REPEATED PATTERN DISCOVERY IN POLYPHONIC MUSIC

Otso Björklund
otso.bjorklund@helsinki.fi

BACKGROUND

Point-set representations of music enable repeated pattern discovery in polyphonic music. Multiple algorithms for repeated pattern discovery using point-sets exist, e.g., SIA, SIATEC [1], and SIAR [2]. The algorithms commonly compute *Maximal Translatable Patterns* (MTP) and optionally also all translated occurrences of MTPs as *Translational Equivalence Classes* (TEC).

AIM

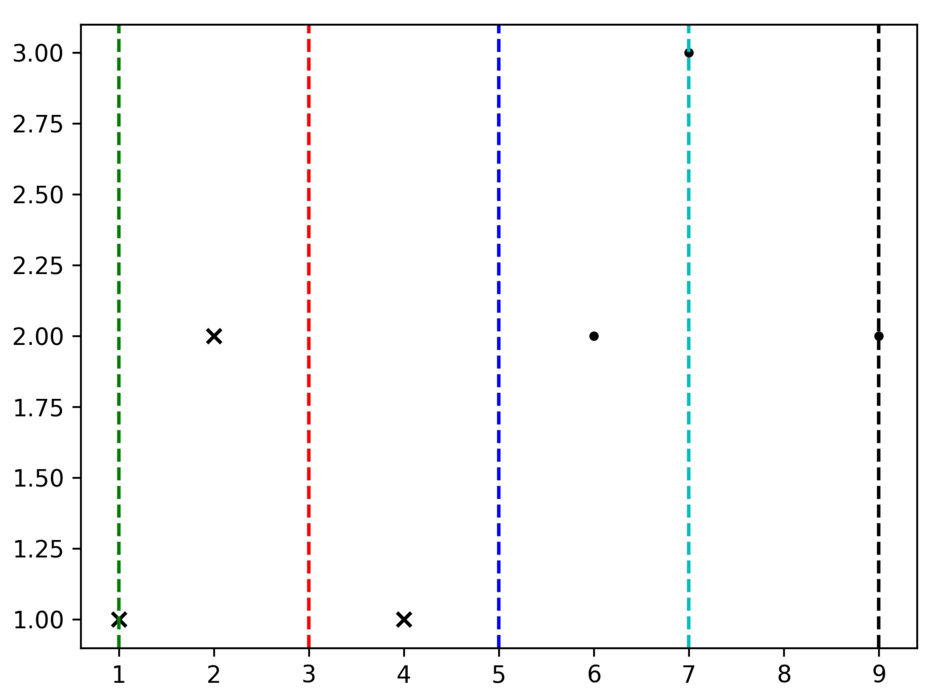
The objective of developing a novel point-set algorithm for repeated pattern discovery is to improve upon the time and space complexity of discovering patterns and all their translated occurrences. With lower worst-case time and space complexity, the algorithm can scale to larger inputs than previous algorithms.

THE SIATEC-C ALGORITHM

The SIATEC-C algorithm takes as its input a point-set and a maximum *inter-onset-interval* (IOI) threshold δ. The algorithm outputs a set of patterns such that the largest IOI between points in any pattern is at most δ. All transposed occurrences of the pattern are output as TECs.

The algorithm discovers patterns by computing MTPs similar to SIA. A sliding window for each point is used to restrict the number of difference vectors that are kept in memory to ensure subquadratic space complexity.

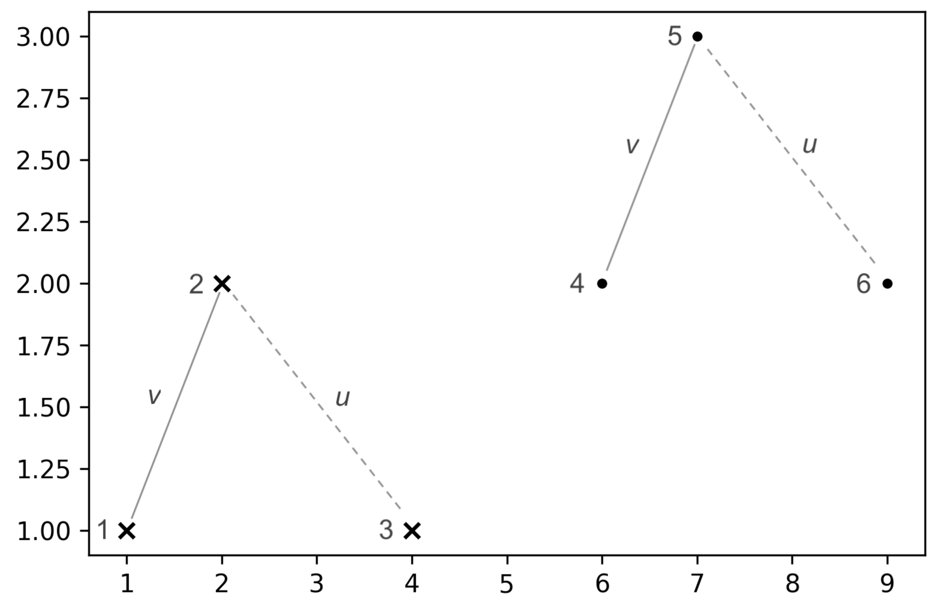
Figure 1: point-specific sliding windows (δ = 2)



The MTPs are split at any gaps exceeding δ and patterns that cover points already covered by a previously discovered larger pattern are discarded.

TECs are computed using an *index structure* that contains an entry for each difference vector between pairs of points, such that, their IOI difference is at most δ. Each entry contains the indices of points that are translatable by the entry's difference (*source indices*) and the indices of points produced by the translation (*target indices*).

Figure 2: index structure (δ = 2)



Difference	Source-target index pairs
(1.0, 1.0)	(1, 2), (4, 5)
(2.0, -1.0)	(2, 3), (5, 6)
(2.0, 1.0)	(3, 4)

Given a point-set D with n points, such that the largest number of point in any time span of length δ is m and the largest MTP in D contains h points, then the worst-case time complexity of SIATEC-C is $O(hn^2 \log nm)$ and the space complexity is $O(nm)$.

RESULTS

The computational performance of SIATEC-C was evaluated on artificial datasets and its precision and recall on JKU-PDD [3].

Figure 3: JKU-PDD est.precision and recall

Algorithm	Corpus	N_{points}	$N_{patterns}$	N_{gt}	P_{est}	R_{est}
SIATEC	monophonic	677.2	30014.8	6.2	0.128	0.679
SIATEC-C ($\delta = 4$)	monophonic	677.2	970.0	6.2	**0.189**	**0.890**
SIAR ($r = 1$)	monophonic	677.2	5365.0	6.2	0.148	0.679
COSIATEC	monophonic	677.2	15.2	6.2	0.136	0.234
SIATECCompress	monophonic	677.2	10.6	6.2	0.124	0.116
COSIATEC-C	monophonic	677.2	28.8	6.2	0.090	0.214
SIATEC-CCompress	monophonic	677.2	21.4	6.2	0.087	0.148
SIATEC	polyphonic	1289.0	59081.8	5.4	0.105	0.690
SIATEC-C ($\delta = 4$)	polyphonic	1289.0	977.6	5.4	**0.131**	**0.775**
SIAR ($r = 1$)	polyphonic	1289.0	12721.4	5.4	0.116	0.635
COSIATEC	polyphonic	1289.0	19.6	5.4	0.091	0.196
SIATECCompress	polyphonic	1289.0	15.8	5.4	0.103	0.121
COSIATEC-C	polyphonic	1289.0	41.2	5.4	0.070	0.161
SIATEC-CCompress	polyphonic	1289.0	24.6	5.4	0.093	0.194

Figure 4: worst-case running times

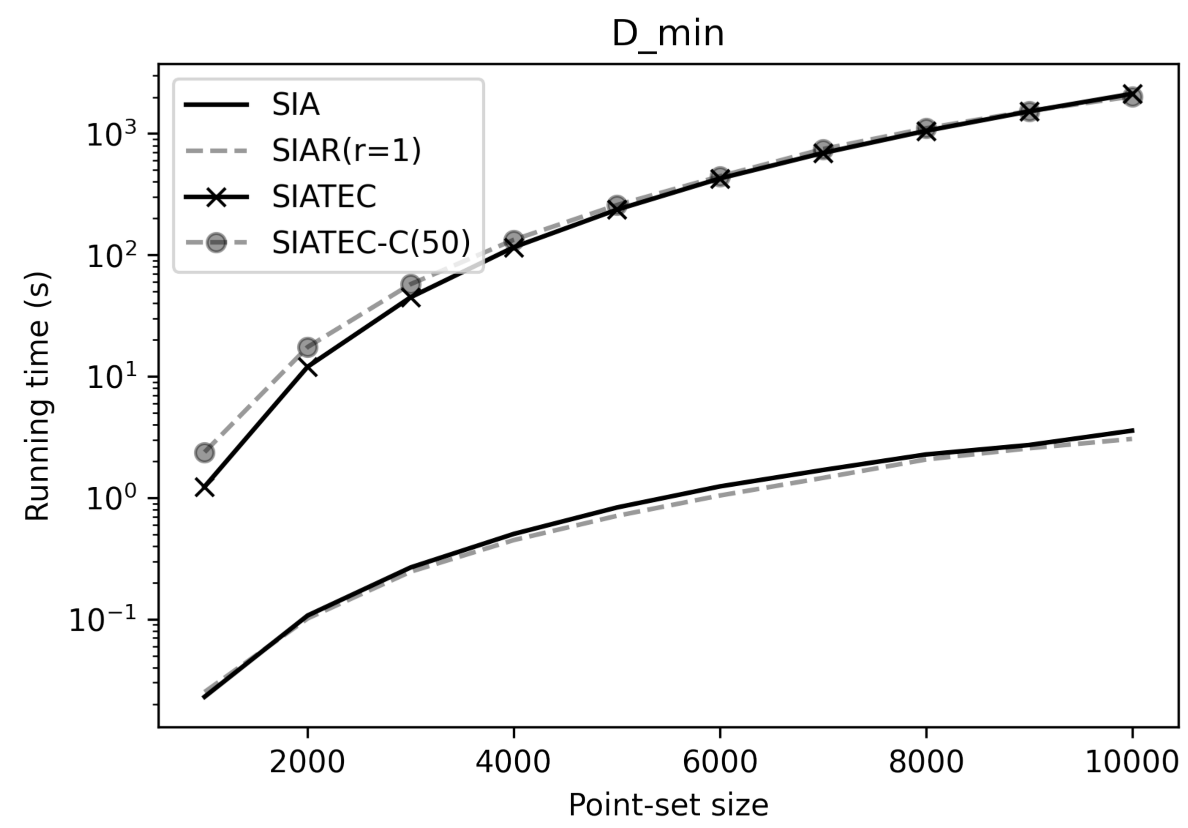
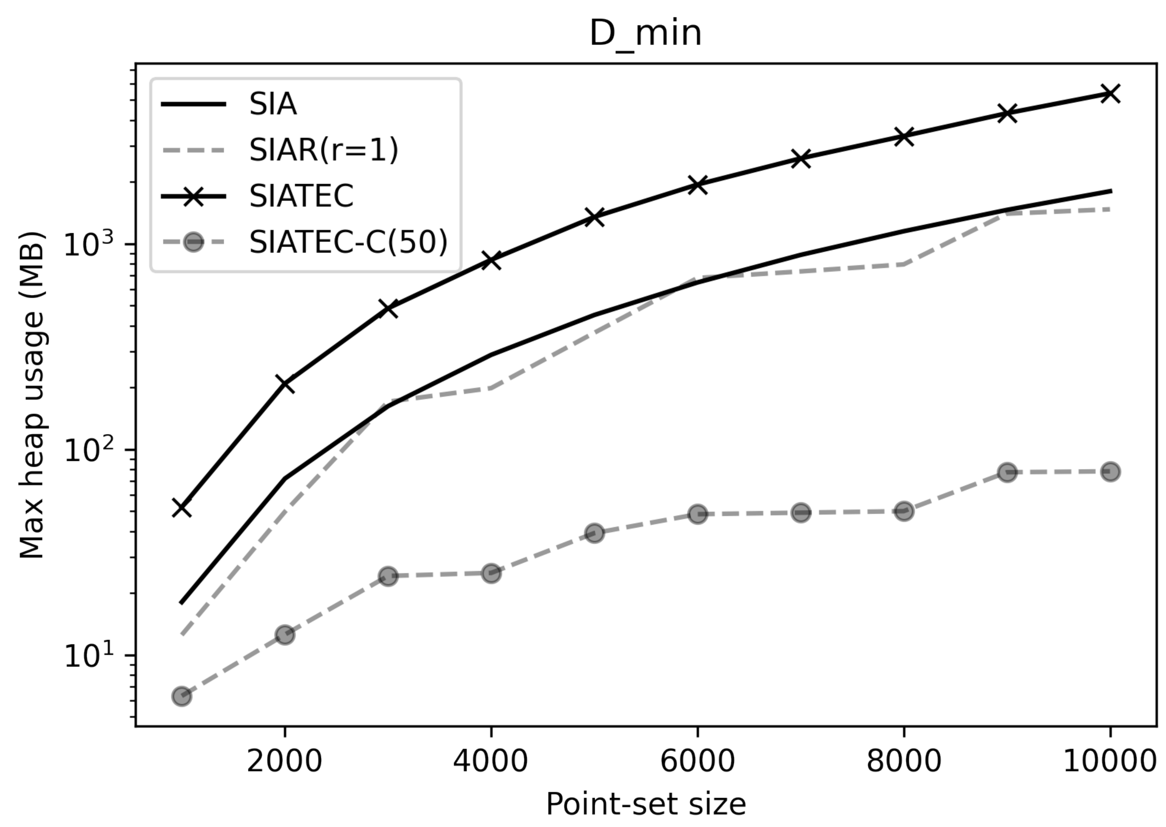


Figure 5: worst-case memory usage



CONCLUSION

SIATEC-C offers a repeated pattern discovery algorithm that can guarantee scaleability to large point-sets.

REFERENCES

[1] D. Meredith, K. Lemström, and G. A. Wiggins, "Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music," *Journal of New Music Research*, vol. 31, no. 4, pp. 321–345, 2002.

[2] T. Collins, Improved methods for pattern discovery in music, with applications in automated stylistic composition. PhD thesis, The Open University, 2011.

[3] T. Collins, "Mirex 2013 competition: Discovery of repeated themes and sections," 2013, accessed: 12 April 2022. [Online]. Available: https://www.music-ir.org/mirex/wiki/2013:Discovery_of_Repeated_Themes_\%26_Sections