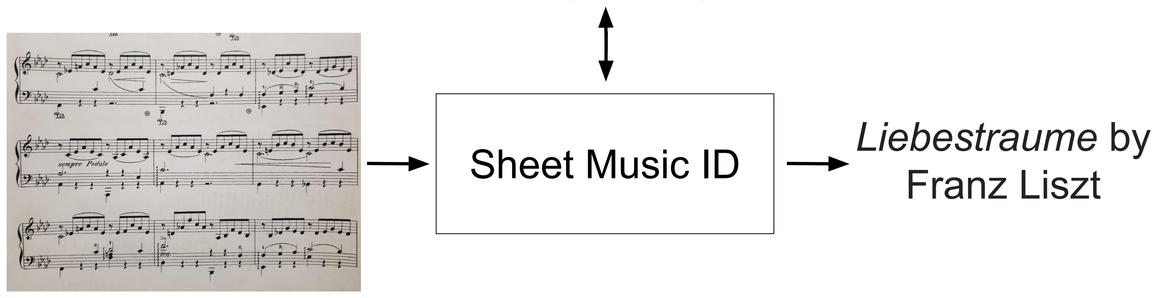
Camera-Based Piano Sheet Music Identification

Daniel Yang, TJ Tsai
Harvey Mudd College

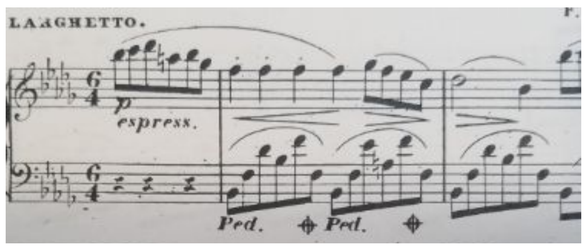
Problem Statement

Given a cell phone image of piano sheet music, identify the piece in IMSLP.

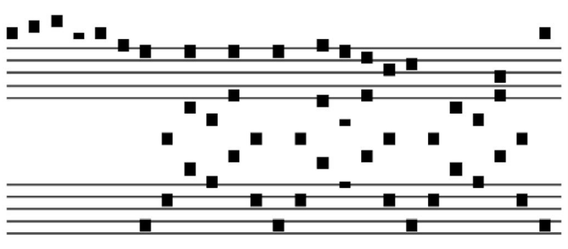


Offline Component

1 Compute Bootleg Score



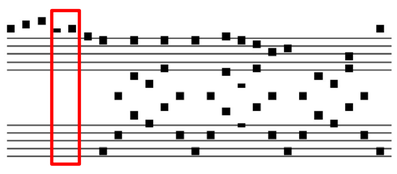
Excerpt of Sheet Music Image



Corresponding Bootleg Score

- mid-level representation that encodes position of filled noteheads [1]
- 62 x N binary matrix
- takes ~1sec/page to compute

2 Construct N-grams



$$\rightarrow x_1, x_2, x_3, \boxed{x_4, x_5}, x_6, \ldots$$

2-Gram

- Represent each bootleg score column as a 64-bit integer
- Every n consecutive integers constitutes an n-gram fingerprint

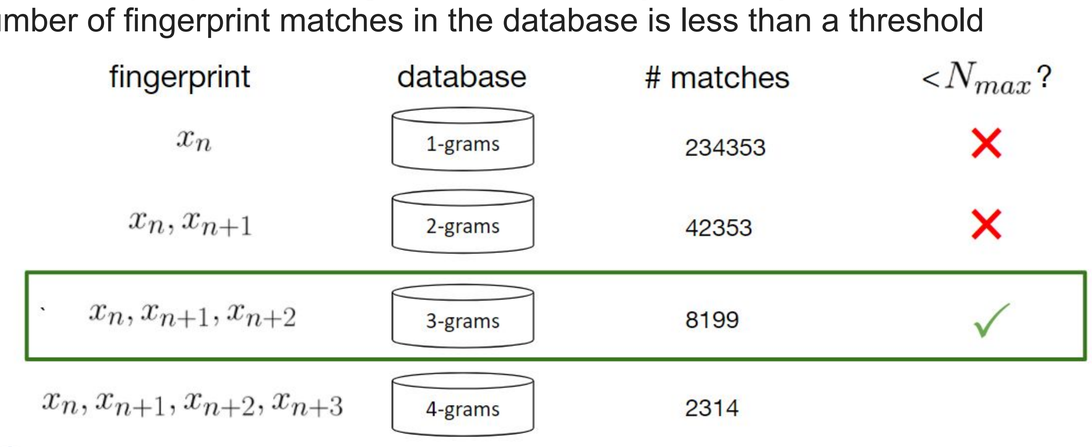
3 Construct Database

- N-gram fingerprints with N=1 to 4 compiled into reverse index
- Keys are fingerprints, values are pieces and offsets where they occur

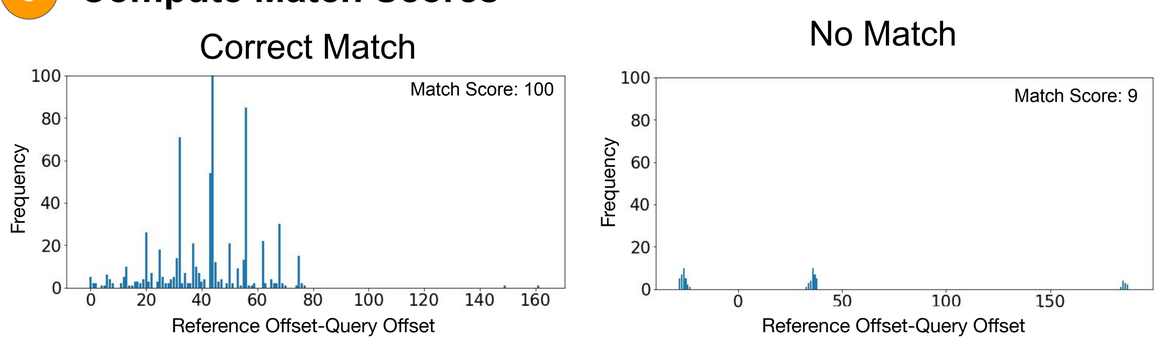
Online Component

4 Construct Dynamic N-grams

Construct a sequence of n-grams where n is dynamically chosen such that the number of fingerprint matches in the database is less than a threshold

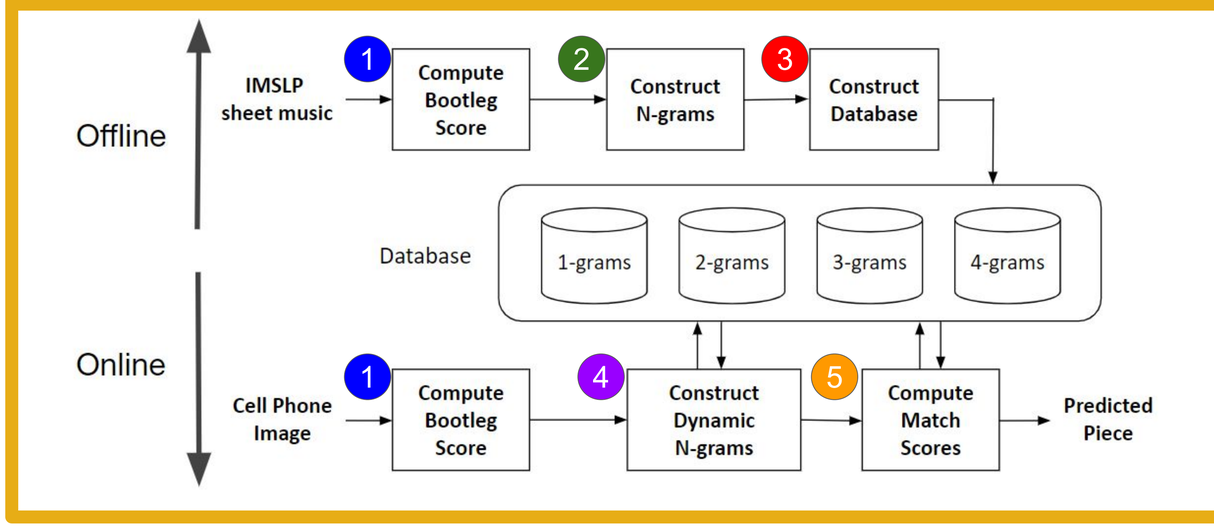


5 Compute Match Scores



Rank database items using histogram of offsets [2]

System Overview



Experimental Setup

Database

- All IMSLP solo piano sheet music
- 29,310 total pieces
- 374,758 total pages

Queries

- Taken from SMR Dataset [3]
- 2000 cell phone pictures of physical sheet music
- Images from 200 piano scores
- Each image shows 1-5 lines of music

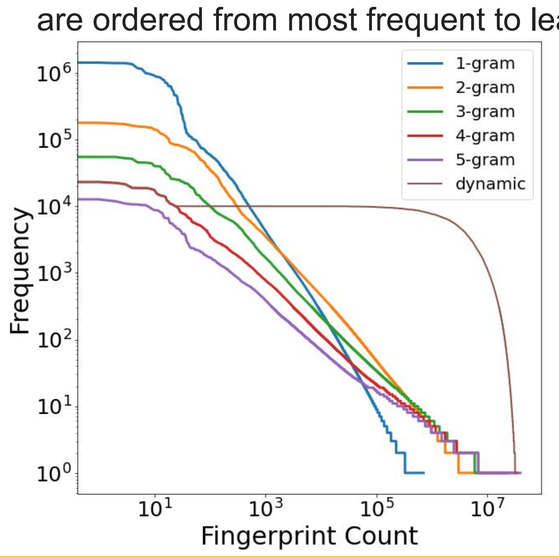
Results

System	MRR	Avg Runtime	Std Runtime
1-gram	.709	21.5s	12.5s
2-gram	.845	2.76s	.36s
3-gram	.808	1.99s	.21s
4-gram	.755	1.23s	.13s
5-gram	.688	1.07s	.08s
dynamic	.853	.98s	.12s

- Fixed n-gram has a trade off between accuracy and runtime
- Dynamic n-gram system has the highest accuracy and fastest runtime

Analysis

Plot shows frequency of each fingerprint in the database, where fingerprints are ordered from most frequent to least frequent.



- Ideal distribution for hashing is flat (uniform)
- For short n-grams, there are some fingerprints that are extremely common in the database. Leads to long runtimes.
- For long n-grams, there are a lot of fingerprints that are extremely rare (occur less than 10 times in IMSLP). Leads to worse accuracy.
- Dynamic n-gram system has a much flatter distribution than the fixed n-gram systems.

References

[1] Yang et al. “MIDI passage retrieval using cell phone pictures of sheet music” in ISMIR 2019.
[2] Wang. “An industrial strength audio search algorithm” in ISMIR 2003.
[3] Tsai et al. “Using Cell Phone Pictures of Sheet Music to Retrieve MIDI Passages” in IEEE Transactions on Multimedia 2020.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Large-scale computations were performed with XSEDE Bridges at the Pittsburgh Supercomputing Center through allocation TG-IRI190019.