Identification of Printed Music

Aditya Majumdar (am3713)
Andrew Mercer-Taylor (ajm2209)
Robert Ying (ry2242)

Computer Vision & Machine Learning for Mobile Platforms
Spring 2014, Prof. Belhumeur

Motivation

- Difficult to identify printed sheet music without significant background knowledge
- Often only first page of piece contains musical metadata
  - This metadata may be insufficient to identify a piece; J.S. Bach called more than 150 pieces “Prelude”
- Especially problematic for those learning a new piece or instrument

Project

- Android
- Take pictures of musical score/sheet music
  - Segment them, find interesting features
- Compare against database of sheet music
- Use heuristic to decide how closely the features correspond to database images
- Return piece metadata and a link to a performance on Youtube

Database

- Downloaded from IMSLP (International Music Score Library Project)
- Over 150 distinct pieces in latest version, most with multiple editions
  - Represents selected works from J.S. Bach, Claude Debussy, and Joseph Haydn
- Descriptors are precomputed and stored in place of raw images

Implementation

- All image processing occurs on backend
- Automatically set threshold to ensure functionality under various lighting conditions
- Detect ORB features, calculate descriptors
  - ORB is ~350 times faster than SIFT
  - Noise resistant and rotation invariant
- Matching using parallel linear search for nearest neighbors

Challenges: Computational

- Originally used SIFT as a descriptor, but size of data set made storing SIFT prohibitively large, and would require paging onto disk
  - ORB descriptors take less space and work well with the nature of the images
- Takes time to format and populate database
- The computational cost of searching for the best match grows proportionally to the size of the database
  - Optimizations to mitigate this challenge include precomputed descriptors, parallelization, and a reverse matching strategy

Challenges: Accuracy

- Some scans are of very poor quality
  - Unfocused images tend to be matched to unfocused scans, regardless of the music
- User images of pieces are unpredictable in terms of homography and lighting conditions
  - ORB appears resilient to alterations of homography
  - Rotation invariance proves useful; informal user studies indicate that people tend to assume either landscape or portrait mode will work
  - Thresholding works well even with the dramatically nonlinear illumination of a flash

Challenges: Android

- Troubles with flash and autofocusing meant that we would see different results when testing the same piece on different phones
- Refresh was buggy - often would see remnants of previously captured photos on the screen while taking the next picture
- Initially wanted to play the audio file of the recognized piece immediately, but not every piece is on Youtube

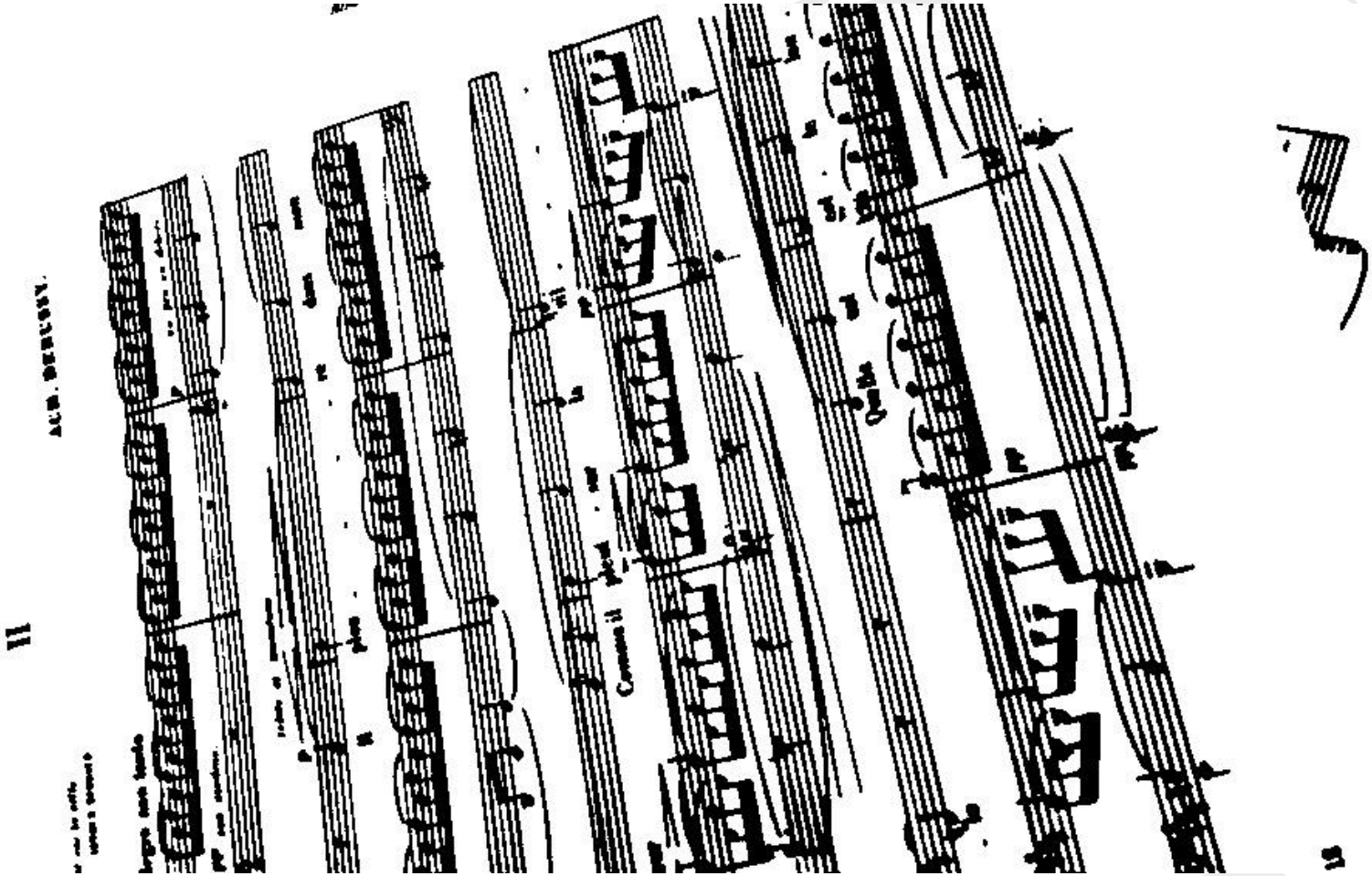
Further Steps

- More composers, pieces in the database
  - This will require a fundamental alteration of the search strategy; an O(n) global database scan is not feasible
- Possible new search approaches:
  - Selecting clusters to search using PCA
  - Template-matching to restrict search space by key signature
- Quantification of performance
  - Initial sanity check matched digitally skewed and distorted scores against the originals, but this is a just crude imitation of real user images

Distribution of Work

- Aditya Majumdar
  - Backend, YouTube integration, frontend assistance, user studies
- Andrew Mercer-Taylor
  - Experimental validation of SIFT method, Android frontend, client communication workflow
- Robert Ying
  - Backend, image processing, parallel image search implementation
- All
  - Idea formation, testing, debugging, acquiring music, and researching solutions

Example Result



title: Ariettes oublies (Debussy Claude)
link: https://www.youtube.com/watch?v=sCgiCaG_dC4

Demo