Probabilistic Music-Symbols Spotting in Handwritten Scores

Jorge Calvo-Zaragoza, Alejandro H. Toselli, Enrique Vidal
PRHLT Research Center, Universitat Politècnica de València
{jcalvo,ahector,evidal}@prhlt.upv.es

1. Introduction

- Direct content-based search on music score images
- Approach based on an efficient probabilistic indexing
- Work focused on handwritten Mensural notation
- Promising alternative to traditional content-based music search

2. Base recognizer

- Input staff-section image
- Recognition based on hidden Markov models and k-grams
- Decoding with Viterbi search algorithm
- N-best hypotheses represented by a Symbol Lattice (SL)

3. Framework: single-symbol spotting

- $\mathcal{R}$ denotes whether a staff image $\mathbf{x}$ is relevant for the music symbol v
- We compute $P(\mathcal{R} \mid \mathrm{v}, \mathbf{x})$ following a posteriorgram-based approach: maximum of the *frame-level symbol score*

$$P(\mathcal{R} \mid \mathrm{v}, \mathbf{x}) \approx \sum_{\mathbf{s} \in \Sigma^* \, \mathrm{v} \, \Sigma^*} P(\mathbf{s} \mid \mathbf{x}) \approx \max_{1 \le i \le M} S(\mathrm{v}, \mathbf{x}, i)$$

- Score $S(\mathrm{v}, \mathbf{x}, i)$ computed from SL:

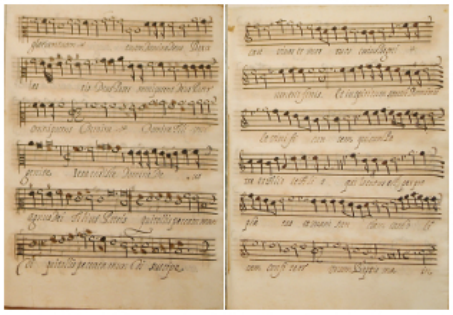
$$S(\mathrm{v}, \mathbf{x}, i) \stackrel{\text{def}}{=} \sum_{e=(q',q):\mathrm{v}=\omega(e), t(q')<i\le t(q)} \frac{\alpha(q')l(q',q)\beta(q)}{\beta(q_I)}$$

4. Framework: multiple-symbols indexing

- Queries formulated as boolean combinations of several music symbols, $\mathrm{v}_1, ..., \mathrm{v}_K$
  - AND: $P(\mathcal{R}_1 \wedge \mathcal{R}_2 \cdots \wedge \mathcal{R}_K) \approx \min(P(\mathcal{R}_1), P(\mathcal{R}_2), \ldots, P(\mathcal{R}_K))$
  - OR: $P(\mathcal{R}_1 \vee \mathcal{R}_2 \cdots \vee \mathcal{R}_K) \approx \max(P(\mathcal{R}_1), P(\mathcal{R}_2), \ldots, P(\mathcal{R}_K))$
  - NOT: $P(\neg B) = 1 - P(B)$

5. Corpora

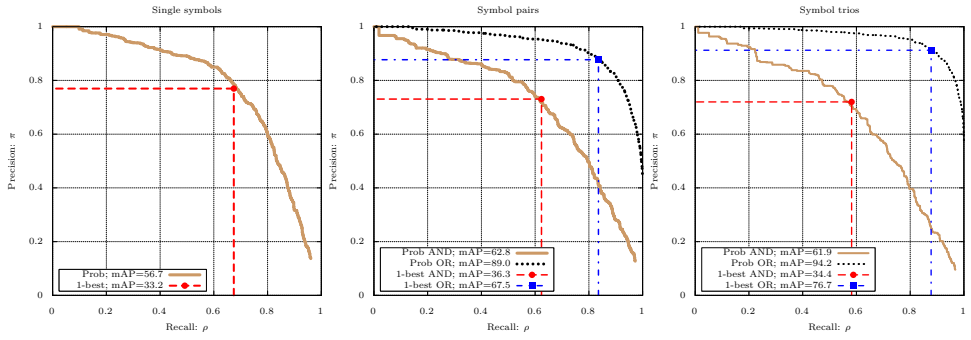
Handwritten mensural notation from Capitan dataset



- Segmentated into isolated staves

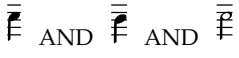
6. Results

- Query set based on the test partition
  - 115 single symbol, 664 symbol pairs, 819 symbol trios
- Comparison with transcription-based retrieval (1-best)

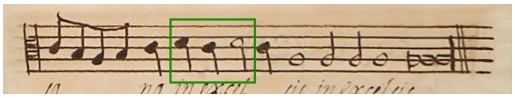


7. Example

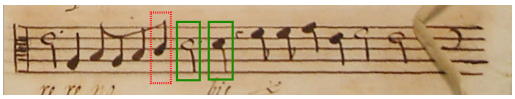
- Symbolic query



- True positive section for thresholds below 0.91



- False positive section for thresholds below 0.97



8. Conclusions

- First step towards content-based music search from untranscribed images
- Efficient approach based on a probabilistic single-symbol index
- Extension to multiple-symbol search related by boolean operations
- Improvement around 20% with respect to the transcription-based search
- Promising future work
  - Replace base recognizer with Recurrent Neural Networks
  - Extend the formulation for sequential search
  - Music-directed search (like melodies or intervals)