MIREX 2015 QBSH TASK: TENCENT BESTIMAGE'S SOLUTION

Bilei Zhu
Tencent Technology (Shanghai) Co.,Ltd
bileizhu@tencent.com

Hong Liu
Xiamen University
lynnliu0207@163.com

ABSTRACT

This extended abstract describes our submission to the task of Query by Singing/Humming (QBSH), submitted to MRIEX 2015. Our system take advantage of the difference of different parts of each query, then divide each query by two different methods, and use each segment phrase to do separately humming search. For each search, EMD-based and DTW-based retrieval scheme are kindly used here. At last, we define a fusion function to combine all the segment phrases searching lists, which can obtain more accurate matching result.

1. INTRODUCTION

Query by Singing/Humming is a well known content-based searching method in music information retrieval. A QBSH system allows user to find a song even if he merely know the tune from parts of the song. User can simply hums a piece of song by a microphone, and the system retrieve the database of songs which contain the similar tune and return a rank list about the search result. At last, the needed song will be get in this list by listening all.

There are two sub-tasks in Query by Humming/Singing Evaluation this year. Task 1 is the classic QBSH evaluation, which aim at finding the ground-truth midi files in song database from a user's singing or humming query with the .wav format. The database contain two parts: ground-truth and noise midi files, comprised of ground-truth midis from MIR-QBSH corpus(48) and IOACAS corpus(106), along with a cleaned version of Essen Databases. As the evaluation, Top-10 hit rate is calculated by top-10 candidate list get from system output. The task 2 is a variants QBSH evaluation, which is based on Prof. Downie's idea that queries are variants of ground-truth midi, and it becomes more important since user-contributed humming is an important part of the song database to be searched, as evidenced by QBSH search service [1] . It has to be mentioned that our submission is dedicated to classical task.

[1] www.midomi.com

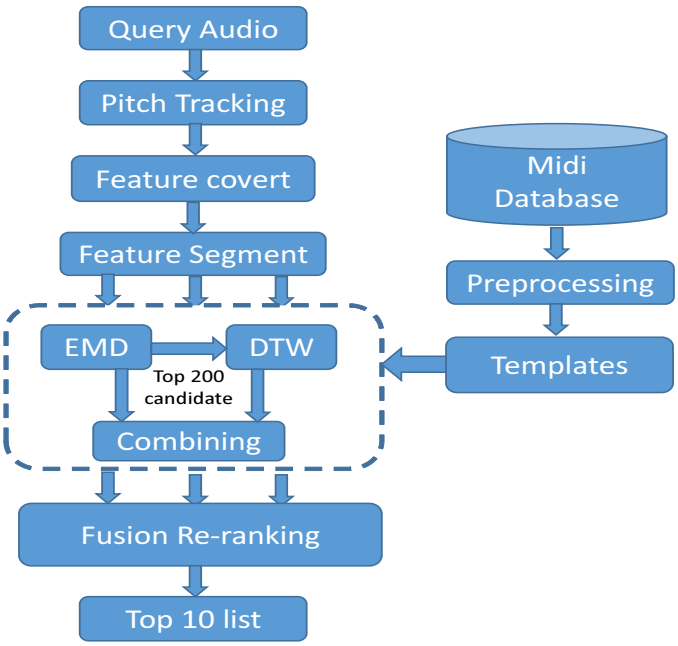

Figure 1. The FrameWork of Our System.

2. SYSTEM DESCRIPTION

Figure 1 shows the framework of the proposed system. Our system consist of four modules: 1) database build module takes music score to generate song template database; 2) pitch tacker module, taking user's singing/humming as input query to generate pitch features and note features. 3) features segments and template matching, computing the similarity between template and each query segment, and generating a ranked list sorted by similarity. 4) segment fusion reranking, computing the final ranking score according to all the ranked lists get by module 3.

2.1 Pitch Tracking and Segment

The accuracy of pitch tracking play an important role in QBSH system. In our pitch tracking module, we use the pYin [1, 2] method to extract the pitch feature of human query, which was proposed by Mauch in 2014. The pYin method adds a HMM-based F0 tracking stage in order to find a "smooth" path through the fundamental frequency candidates obtained by Yin. And we also adopt a five-point median filter to generate smoother pitch sequence, for solving traditional pitch errors, such as small pitch fluctuation.

In our system, we adopt the segment schemes which

divide each query to many short music phrases. For MIR-QBSH corpus, the query beginning hold as the phrase beginning, and setting the end of phrase at silent position in original query sequence. For IOACAS corpus, we separate the each query by the silent phrase.

2.2 Template Matching

Classical sequential matching algorithms, *e.g.* linear Scaling (LS) [3], Recursive Alignment (RA) [4] and Dynamic Time Warping (DTW) [5], have been used in many different QBSH systems. DTW outperforms all the other matching algorithms, but it cannot be used in our template matching module due to its high complexity. EMD [6], for note-based matching, is a good choice to be used as a filtering scheme for reducing overall searching time. That is to say, we filter out most wrong candidates by EMD matching and use DTW matching for accurate search.

2.3 Segment Fusion Reranking

Although we notice that the rank list searching by some phrases can get expected result for evaluation of QBSH system, the difference among top-rank candidates of all the lists are significant. Intuitively, combination of all the ranking list is a good strategy to enhance the searching performance. Suppose a set of ranking lists $L = \{l^1, l^2, ..., l^n\}$, where n is the total number of segments, and $l^i = \{l^i_1, l^i_2, ..., l^i_N\}$ is the ranking list of independent segment, l^i_j is the index of Song ID in database and N is the number of Song database. For the final ranking score scr_i of i-th song, a fusion function is define as follow,

$$src_i = \sum_{k=1}^{n} \frac{1}{l_i^k}. \quad (1)$$

Then the final top-10 rank list can be generated by sorting the score list $Src = \{src_1, src_2, ..., src_N\}$.

3. REFERENCES

[1] Molina, Emilio and Tardón, Lorenzo J and Barbancho, Isabel and Barbancho, Ana M: "The importance of F0 tracking in query-by-singing-humming," *Proceedings of the International Symposium on Music Information Retrieval*, pp.227–282, 2014.

[2] M. Mauch, and S. Dixon: "pYIN: A fundamental frequency estimator using probabilistic threshold distributions," *Proceedings of ICASSP*, 2014.

[3] J.R. Jang, C. Hsu, and H. Lee: "Continuous HMM and its enhancement for singing/humming query retrieval," *Proceedings of International Society for Music Information Retrieval*, pp.546-551, 2005.

[4] X. Wu, M. Li, J. Liu, J. Yang, and Y. Yan: "A top-down approach to melody match in pitch contour for query by hummingl" *Proceedings of International Conference of Chinese Spoken Language Processing,* pp.546-551, 2005.

[5] C. Myers, L. Rabiner, and A. Rosenberg: "Performance tradeoffs in dynamic time warping algorithms for iso- lated word recognition" *IEEE Transactions on Acoustics, Speech and Signal Processing,* 28(6):623?635, 1980.

[6] Typke R, Giannopoulos P, Veltkamp R C: "Using transportation distances for measuring melodic similarity" *Proceedings of International Society for Music Information Retrieval,* 2003.