MIREX 2006

Computing Statistical Spectrum Descriptors for Audio Music Similarity and Retrieval

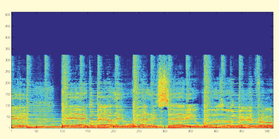
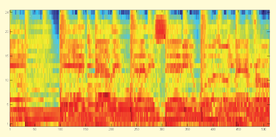
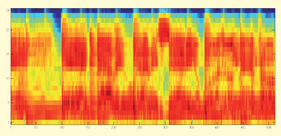
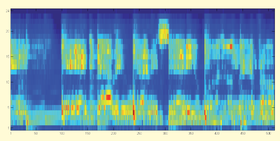
Thomas Lidy, Andreas Rauber
Vienna University of Technology
Department of Software Technology and Interactive Systems
Vienna, Austria
{lidy, rauber}@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/mir

Statistical Spectrum Descriptors

We used a new Java implementation of the SSD feature set which has been also used (partly) in the MIREX 2005 Audio Genre Classification. The Java program is more robust than the previous Matlab code and supports a number of convenient features. It also enables extraction of Rhythm Patterns and Rhythm Histogram features. As the computational cost for extracting SSD is lower than for both the two other feature sets and as pre-liminary tests delivered reasonable results for SSD features employed to retrieval with similarity rankings, we based our submission solely on the SSD features. Also, the dimensionality is lower than the one of Rhythm Patterns, which reduces the cost for distance calculations.

1. conversion to mono + segmentation (5.9 second segments)
2. Fourier Transform (STFT) -> Spectrogram
3. apply Bark scale -> 24 critical bands
4. convert to decibel scale
5. compute loudness level -> Phon scale
6. compute specific loudness sensation -> Sone scale
7. compute statistics per critical band in order to describe fluctuations: mean, median, variance, skewness, kurtosis, min- and max-value

(1)

(2)

(3)

(5)

(6)


A feature vector for an audio file is represented by the median of the SSD features of its segments.

Ref.: T. Lidy and A. Rauber. Evaluation Of Feature Extractors And Psycho-acoustic Transformations For Music Genre Classification. Proc. Intl. Conference on Music Information Retrieval, London, UK, Sep. 2005

Distance Matrix Computation

As the the implementation of similarity ranking and according distance measures in our Java software has not yet been completed, we ran the distance matrix calculation in Matlab. The distance matrix is calculated from the SSD features using the cityblock metric.

Audio Cover Song Identification

Task

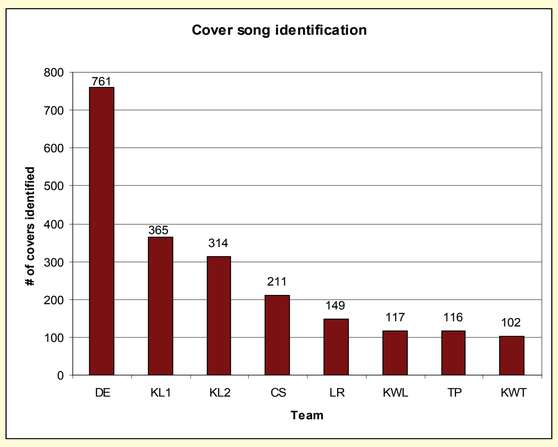
- 30 cover songs of a variety of genres
- 11 versions each
- i.e. 330 audio files
- 30 queries with cover songs
- list of 10 returned items is examined for the presence of 10 cover songs

Participants

CS = Christian Sailer (1)
DE = Daniel P.W. Ellis (1)
KL1/2 = Kyogu Lee 1 (1)
KWL = Kris West (Likely) (2)
KWT = Kris West (Trans) (2)
LR = Thomas Lidy & Andreas Rauber (2)
TP = Tim Pohle (2)

(1) submissions specifically designed to detect cover song variants.
(2) submissions not-specifically designed to detect cover song variants, but for Audio Music Similarity and Retrieval

Results



Friedman Test (p=0.05) against Mean reciprocal rank (the reciprocal of the rank of the first correctly identified cover for each query (1/rank)) showed, that DE is significantly better than the other algorithms, while there is no significant difference between the remaining algorithms.

Audio Music Similarity and Retrieval

Database

- large scale music similarity evaluation
- 5000 audio files (22kHz, mono, 16bit)
- maximum of 20 tracks per artist
- minimum of 50 tracks per labelled genre
- genres: Jazz, Rap & Hip Hop, Latin, Rock, R&B, Reggae, Country, New Age, Electronica & Dance
- contains the 330 songs from the cover song task

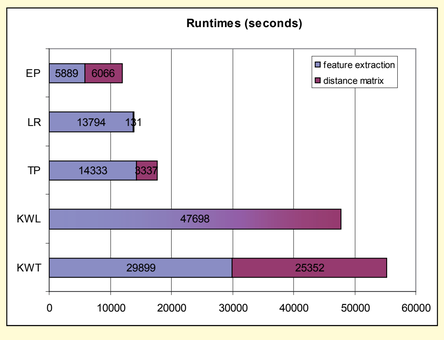
Task

apply feature extraction for audio similarity / retrieval and compute distance matrix between all 5000 songs, for:
1) Human Evaluation of ranked lists
2) Evaluation of Objective Statistics

Participants

EP = Elias Pampalk
TP = Tim Pohle
VS = Vitor Soares
LR = Thomas Lidy & Andreas Rauber
KWT = Kris West (Trans)
KWL = Kris West (Likely)

Runtimes



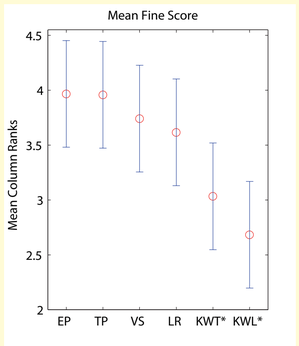
Linux (CentOS)
Dual AMD Opteron 64
1.6 GHz
4 GB RAM

Human Evaluation Results

60 randomly selected queries
~ 20 human evaluators
7-8 queries per evaluator
3 evaluations per query/candidate pair
two evaluation scales:
- broad scale: very similar - somewhat similar - not similar
- fine scale: value between 0 and 10 (10 = best)

Measures:
(All measures normalized to the range 0 to 1)
Fine = Sum of fine-grained human similarity decisions (0-10).
PSum = Sum of human broad similarity decisions: NS=0, SS=1, VS=2.
WCsum = 'World Cup' scoring: NS=0, SS=1, VS=3 (rewards Very Similar).
SDsum = 'Stephen Downie' scoring: NS=0, SS=1, VS=4 (strongly rewards VS).
Greater0 = NS=0, SS=1, VS=1 (binary relevance judgement).
Greater1 = NS=0, SS=0, VS=1 (binary relevance judgement using only VS).

	EP	TP	VS	LR	KWT	KWL
Fine	0,430	0,423	0,404	0,393	0,372	0,339
Psum	0,425	0,411	0,388	0,374	0,349	0,313
Wcsum	0,358	0,340	0,323	0,306	0,280	0,248
Sdsum	0,324	0,305	0,290	0,271	0,246	0,216
Greater0	0,627	0,623	0,586	0,579	0,557	0,509
Greater1	0,223	0,199	0,191	0,169	0,142	0,118

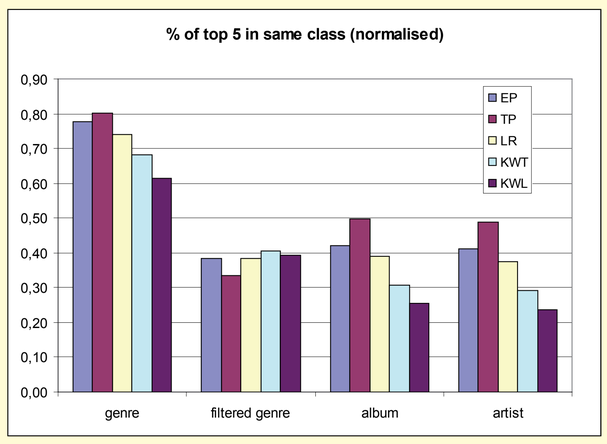


Plot of Friedman Test results:

Friedman Test with Multiple Comparisons Results (p=0.05) showed that there **no significant differences** in the results of all algorithms, except for the KWL implementation compared to EP, TP or VS.

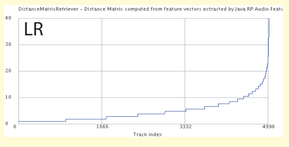
Objective Statistics Results

- Average % of Genre, Artist and Album matches in the top 5, 10, 20 & 50 results
- Precision at 5, 10, 20 & 50
- Average % of Genre matches in the top 5, 10, 20 & 50 results after artist filtering of results
- Always similar - Maximum # times a file was in the top 5, 10, 20 & 50 results
- % of song triplets where triangular inequality holds
- etc.

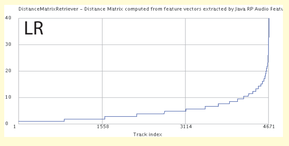


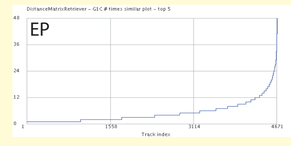
	EP	TP	LR	KWT	KWL
genre	0,78	0,80	0,74	0,68	0,61
filtered genre	0,38	0,33	0,38	0,41	0,39
album	0,42	0,50	0,39	0,31	0,25
artist	0,41	0,49	0,37	0,29	0,24

Plot of the "number of times similar curve" - plot of song number vs. number of times it appeared in a top 5 list with songs sorted according to number times it appeared in a top 5 list. Systems with a sharp rise at the end of this plot have "hubs" (i.e. always similar songs), while a long 'zero' tail shows many never similar results.



with cover songs filtered:





Part of this work was supported by the EU in the 6th Framework Program, IST, through the MUSCLE NoE, contract 507752.

Presented at the 7th International Conference on Music Information Retrieval, Victoria, Canada, 8 - 12 October 2006.