Informed Feature Representations for Music and Motion

Meinard Müller

Lorentz Workshop
Music Similarity: Concepts, Cognition and Computation

Thanks
- Sebastian Ewert
- Peter Grosche
- Andreas Baak
- Tido Röder

Overview
- Audio Features based on Chroma Information Application: Audio Matching
- Motion Features based on Geometric Relations Application: Motion Retrieval
- Audio Features based on Tempo Information Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema Application: Data-Driven Motion Reconstruction

Meinard Müller
- 2007 Habilitation, Bonn
- 2007 MPI Informatik, Saarbrücken Senior Researcher
  Music Processing & Motion Processing
- 2012 W3-Professur, AudioLabs Erlangen
  Semantic Audio Processing

Music and Motion

Overview
- Audio Features based on Chroma Information Application: Audio Matching
- Motion Features based on Geometric Relations Application: Motion Retrieval
- Audio Features based on Tempo Information Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema Application: Data-Driven Motion Reconstruction
Chroma-based Audio Features

- Very popular in music signal processing
- Based equal-tempered scale of Western music
- Captures information related to harmony
- Robust to variations in instrumentation or timbre

Example: Chromatic scale

Pitch (MIDI note number)

C8: 4186 Hz
C7: 2093 Hz
C6: 1046 Hz
C5: 523 Hz
C4: 261 Hz
C3: 131 Hz

Intensity (dB)

Log-frequency spectrogram

Time (seconds)
Chroma-based Audio Features

Example: Chromatic scale

Log-frequency spectrogram

Chroma representation (normalized, Euclidean)

Motivation: Audio Matching

Four occurrences of the main theme

Enhancing Chroma Features

- Making chroma features more robust to changes in timbre
- Combine ideas of speech and music processing
- Usage of audio matching framework for evaluating the quality of obtained audio features

M. Müller and S. Ewert
How to make chroma features more robust to timbre changes?

Idea: Discard timbre-related information

How to make chroma features more robust to timbre changes?

Idea: Discard lower MFCCs to achieve timbre invariance
Enhancing Timbre Invariance

**Steps:**
1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Discard lower coefficients $[1:n-1]$

Enhancing Timbre Invariance

**Steps:**
1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients $[n:120]$
5. Inverse DCT

Enhancing Timbre Invariance

**Steps:**
1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients $[n:120]$

Enhancing Timbre Invariance

**Steps:**
1. Log-frequency spectrogram
Enhancing Timbre Invariance

Steps:
1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients \([n:120]\)
5. Inverse DCT
6. Chroma & Normalization

Chroma versus CRP

Shostakovich Waltz

Audio Analysis

Idea:
Use “Audio Matching” for analyzing and understanding audio & feature properties:
- Relative comparison
- Compact
- Intuitive
- Quantitative evaluation
Audio Analysis

**Query:** Shostakovich, Waltz (Yablonsky)

**Expected matching positions (should have local minima)**

---

Audio Analysis

**Idea:**
- Use matching curve for analyzing feature properties

---

Audio Analysis

**Standard Chroma (Chroma Pitch)**

**Query:** Free in you / Indigo Girls (1. occurrence)

**Expected matching positions (should have local minima)**

---

Quality: Audio Matching

**Query:** Free in you / Indigo Girls (1. occurrence)

**Idea:**
- Use matching curve for analyzing feature properties
- **Example:** Chroma feature of higher timbre invariance

---

Quality: Audio Matching

**Query:** Free in you / Indigo Girls (1. occurrence)

**Chroma Toolbox**

- There are many ways to implement chroma features
- Properties may differ significantly
- Appropriateness depends on respective application

- MATLAB implementations for various chroma variants
Overview

- Audio Features based on Chroma Information
  Application: Audio Matching

- Motion Features based on Geometric Relations
  Application: Motion Retrieval

- Audio Features based on Tempo Information
  Application: Music Segmentation

- Depth Image Features based on Geodesic Extrema
  Application: Data-Driven Motion Reconstruction

Motion Capture Data

- 3D representations of motions
- Computer animation
- Sports
- Gait analysis

Motion Capture Data

Optical System

Motion Retrieval

- $D = \text{MoCap database}$
- $Q = \text{query motion clip}$
- **Goal**: find all motion clips in $D$ similar to $Q$
Motion Retrieval

- **Numerical similarity vs. logical similarity**
- Logically related motions may exhibit significant spatio-temporal variations

Relational Features

- Exploit knowledge of kinematic chain
- Express geometric relations of body parts
- Robust to motion variations

Meinard Müller, Tido Röder, and Michael Clausen

*Efficient content-based retrieval of motion capture data.*


Meinard Müller and Tido Röder

*Motion templates for automatic classification and retrieval of motion capture data.*


Relational Features

- Right knee bent?
- Right foot fast?
- Right hand moving upwards?

Motion Templates (MT)
Motion Templates (MT)

Temporal alignment

Superimpose templates

Compute average

Average template
Motion Templates (MT)

Quantized template

- Gray areas indicate inconsistencies / variations
- Achieve invariance by disregarding gray areas

MT-based Motion Retrieval

MT-based Motion Retrieval: Jumping Jack

Features

Time (seconds)
MT-based Motion Retrieval: Jumping Jack

MT-based Motion Retrieval: Elbow-To-Knee

MT-based Motion Retrieval: Cartwheel

MT-based Motion Retrieval: Throw
MT-based Motion Retrieval: Basketball

Overview
- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Audio Features based on Tempo Information
  Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema
  Application: Data-Driven Motion Reconstruction

MT-based Motion Retrieval: Lie Down Floor

Music Signal Processing

Analysis tasks
- Segmentation
- Structure analysis
- Genre classification
- Cover song identification
- Music synchronization
- …
Music Signal Processing

Analysis tasks
- Segmentation
- Structure analysis
- Genre classification
- Cover song identification
- Music synchronization
- ...

Audio features
- Musically meaningful
- Semantically expressive
- Robust to deviations
- Low dimensionality
- ...

Mid-Level Representations

<table>
<thead>
<tr>
<th>Musical Aspect</th>
<th>Features</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timbre</td>
<td>MFCC features</td>
<td>10 - 15</td>
</tr>
<tr>
<td>Harmony</td>
<td>Pitch features</td>
<td>60 - 120</td>
</tr>
<tr>
<td>Harmony</td>
<td>Chroma features</td>
<td>12</td>
</tr>
<tr>
<td>Tempo</td>
<td>Tempogram</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>Tempo</td>
<td>Cyclic tempogram</td>
<td>10 - 30</td>
</tr>
</tbody>
</table>

Peter Grosche, Meinard Müller, and Frank Kurth

Cyclic tempogram – a mid-level tempo representation for music signals.
Novelty Curve

Steps:
1. Spectrogram

Novelty Curve

Steps:
1. Spectrogram
2. Log compression

Novelty Curve

Steps:
1. Spectrogram
2. Log compression
3. Differentiation

Novelty Curve

Steps:
1. Spectrogram
2. Log compression
3. Differentiation
4. Accumulation

Novelty Curve

Steps:
1. Spectrogram
2. Log compression
3. Differentiation
4. Accumulation
5. Normalization
Tempogram

Short-time Fourier analysis

Log-Scale Tempogram
Cyclic Tempogram

Cyclic projection
Relative to tempo class [...,30,60,120,240,480,...]

Quantization: 60 tempo bins

Quantization: 30 tempo bins

Quantization: 15 tempo bins

Audio Segmentation

Example: Brahms Hungarian Dance No. 5

Example: Zager & Evans: In the year 2525
Audio Segmentation

Example: Beethoven Pathétique

Overview

- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Audio Features based on Tempo Information
  Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema
  Application: Data-Driven Motion Reconstruction

Data-Driven Motion Reconstruction

- Goal: Reconstruction of 3D human poses from a depth image sequence
- Data-driven approach using MoCap database
- Depth image features: Geodesic extrema

Andreas Baak, Meinard Müller, Gaurav Bharaj, Hans-Peter Seidel, and Christian Theobalt
A data-driven approach for real-time full body pose reconstruction from a depth camera.

Data-Driven Motion Reconstruction

Input: Depth image          Output: 3D pose

Voting
Database lookup
Local opt.
Previous frame
Input
Output

Data-Driven Motion Reconstruction

Database lookup
Local optimization
Voting scheme
Data-Driven Motion Reconstruction

- Database lookup
- Local optimization
- Voting scheme

Input Output

- Database lookup
- Local optimization
- Voting scheme

Need of motion features for cross-modal comparison

Database Lookup

Depth Image Features
- Point cloud

[Plagemann, Ganapathi, Koller, Thrun, ICRA 2010]

Depth Image Features
- Point cloud
- Graph

[Plagemann, Ganapathi, Koller, Thrun, ICRA 2010]
Depth Image Features
- Point cloud
- Graph
- Distances from root

Observation: First five extrema often correspond to end-effectors and head

Database Lookup

Local Optimization

Voting Scheme
- Combine database lookup & local optimization
- Inherit robustness from database pose
- Inherit accuracy from local optimization pose
- Compare with original raw data pose using a sparse symmetric Hausdorff distance
Voting Scheme
Distance measure (Hausdorff)

Voting Scheme
Distance measure (Hausdorff, symmetric, sparse)

Experiments

Informed Feature Representations
- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Audio Features based on Tempo Information
  Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema
  Application: Data-Driven Motion Reconstruction

Informed Feature Representations
- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Audio Features based on Tempo Information
  Application: Music Segmentation
- Depth Image Features based on Geodesic Extrema
  Application: Data-Driven Motion Reconstruction

- Exploit model assumptions
  - Equal-tempered scale
  - Kinematic chain
- Deal with variances on feature level
  - Enhancing timbre invariance
  - Relational features
  - Quantized motion templates
- Consider requirements for specific application
  - Explicit information often not required
  - Mid-level features

Features with explicit meaning.
Makes subsequent steps more robust and efficient!
Avoid making problem harder as it is.
Conclusions

Selected Publications (Music Processing)


Selected Publications (Motion Processing)