Tutorial Slides

Analysis and Retrieval Techniques for Motion and Music Data

Meinard Müller

Saarland University and MPI Informatik
meinard@mpi-inf.mpg.de

Eurographics 2009

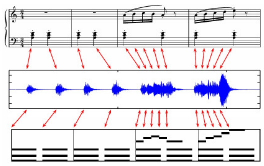
UNIVERSITÄT DES SAARLANDES

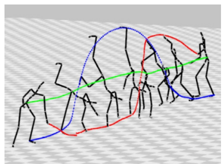
CLUSTER OF EXCELLENCE

max planck institut informatik

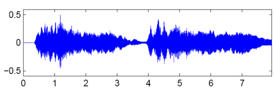
Part 0

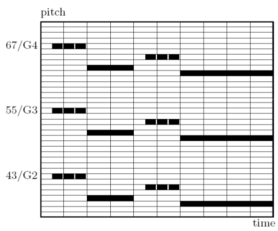
Overview





Music Data





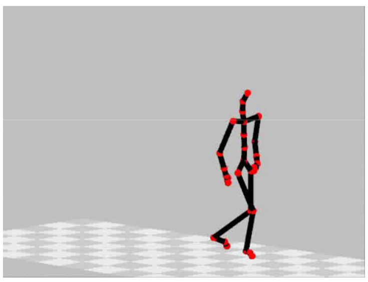
Music Data

Various interpretations – Beethoven‘s Fifth

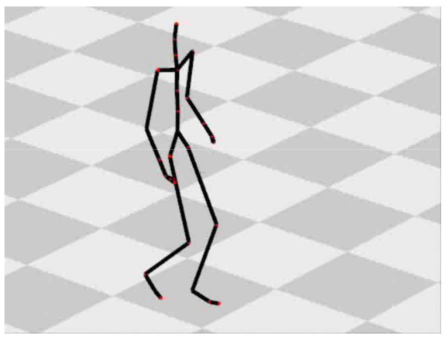
Bernstein	▶
Karajan	▶
Scherbakov (piano)	▶
MIDI (piano)	▶

Motion Capture Data

- Digital 3D representations of motions
- Computer animation
- Sport sciences
- Computer vision



Motion Capture Data



General Tasks

- Automated data organization
- Handling object deformations
- Handling multimodality
- Synchronization (alignment)
- Efficiency

Overview

Part I: Music Synchronization

Part II: Audio Structure Analysis

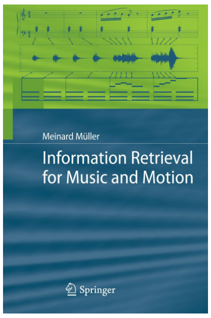
Part III: Audio Matching

Part IV: Motion Retrieval

Bonn University

- Prof. Dr. Michael Clausen
- PD Dr. Frank Kurth
- Dipl.-Inform. Christian Fremerey
- Dipl.-Inform. David Damm
- Dipl.-Inform. Sebastian Ewert
- Dr. Tido Röder

Habilitation



Bonn University

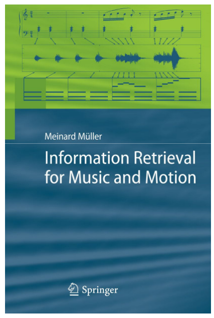
- Prof. Dr. Michael Clausen
- PD Dr. Frank Kurth
- Dipl.-Inform. Christian Fremerey
- Dipl.-Inform. David Damm
- Dipl.-Inform. Sebastian Ewert
- Dr. Tido Röder

Dec. 2007

PhD students

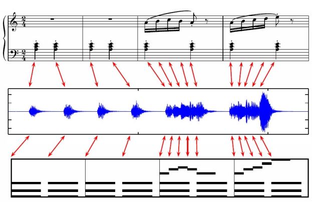
- Dipl.-Inform. Andreas Baak (DFG)
- Dipl.-Math. Verena Konz (MMCI)
- Dipl.-Ing. Peter Grosche (MMCI)
- Dipl.-Inform. Thomas Helten (DFG)

Habilitation

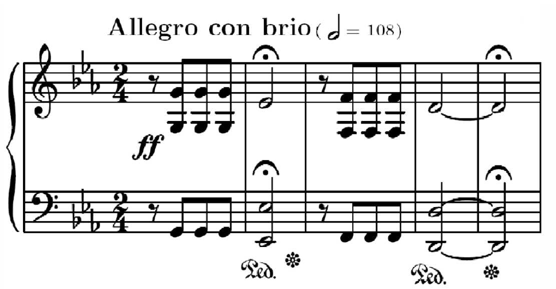


Part I

Music Synchronization



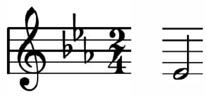
Score Representation



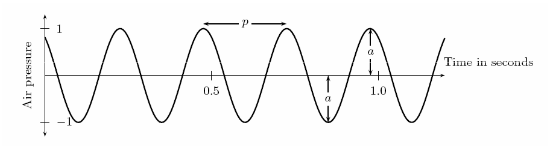
Score Representation: Scanned Image

Score Representation: MusicXML

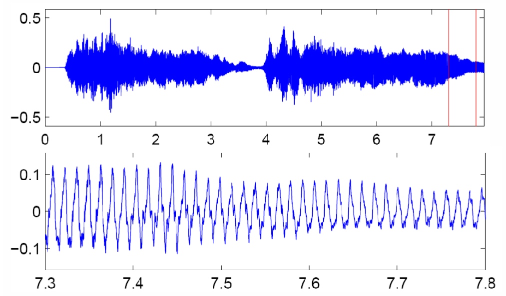
```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```



Audio Representation: Waveform

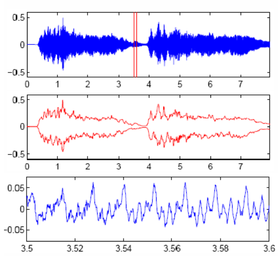


Audio Representation: Waveform

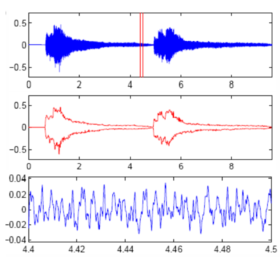


Audio Representation: Waveform

Bernstein (orchestra)



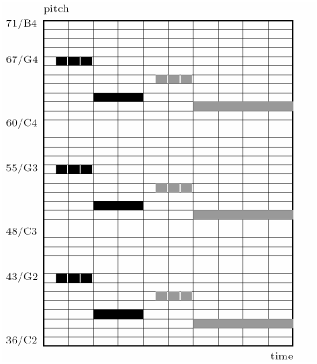
Glen Gould (piano)



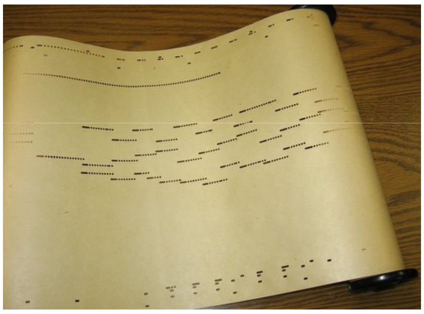
MIDI Representation

Ticks	Message	Ch.	MNN	Vel
60	NOTE ON	1	67	100
0	NOTE ON	2	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	2	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	2	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	2	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	2	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	2	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0

MIDI Representation: Piano Roll



MIDI Representation: Piano Roll



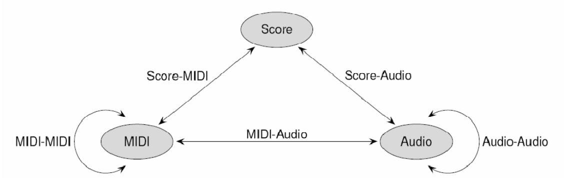
MIDI Representation: Piano Roll

General Goals

- Automated organization of complex and inhomogeneous music collections
- Generation of annotations and cross-links
- Tools and methods for multimodal search, navigation and interaction

Music Information Retrieval (MIR)

Music Synchronization



Schematic view of various synchronization tasks

Music Synchronization

- Turetsky/Ellis (ISMIR 2003)
- Soulez/Rodet/Schwarz (ISMIR 2003)
- Arifi/Clausen/Kurth/Müller (ISMIR 2003)
- Hu/Dannenberg/Tzanetakis (WASPAA 2003)
- Müller/Kurth/Röder (ISMIR 2004)
- Raphael (ISMIR 2004)
- Dixon/Widmer (ISMIR 2005)
- Müller/Mattes/Kurth (ISMIR 2006)
- Dannenberg /Raphael (Special Issue ACM 2006)
- Kurth/Müller/Fremerey/Chang/Clausen (ISMIR 2007)
- Fujihara/Goto (ICASSP 2008)
- Wang/Iskandar/New/Shenoy (IEEE T-ASLP 2008)

Music Synchronization: Audio-Audio

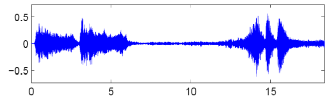
Given: Two different audio recordings of the same underlying piece of music.

Goal: Find for each position in one audio recording the musically corresponding position in the other audio recording.

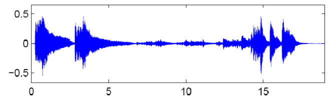
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan



Scherbakov

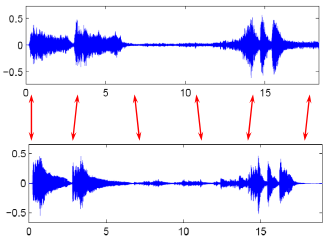


Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan

Scherbakov

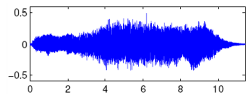


Synchronization: Karajan → Scherbakov

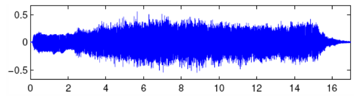
Music Synchronization: Audio-Audio

Bach Toccata

Koopman



Ruebsam

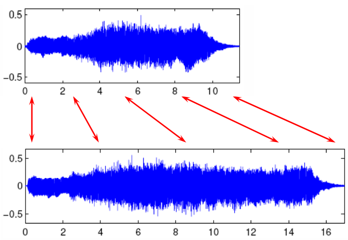


Music Synchronization: Audio-Audio

Bach Toccata

Koopman

Ruebsam



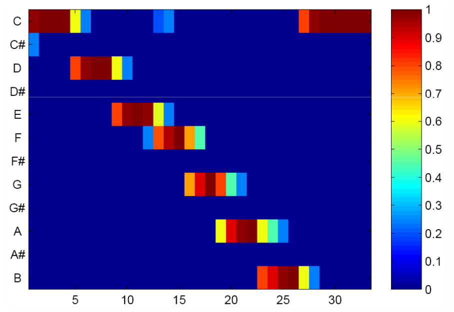
Synchronization: Koopman → Ruebsam

Music Synchronization: Audio-Audio

- Transformation of audio recordings into sequences of feature vectors
  $\rightsquigarrow \; V := (v^1, v^2, \ldots, v^N)$
  $\rightsquigarrow \; W := (w^1, w^2, \ldots, w^M)$
- Fix cost measure c on the feature space
- Compute $N \times M$ cost matrix $C(n,m) := c(v^n, w^m)$
- Compute cost-minimizing warping path from C

Chroma Features

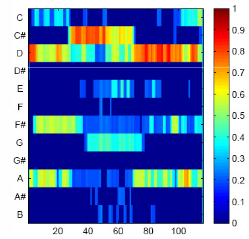
Example: C-Major Scale



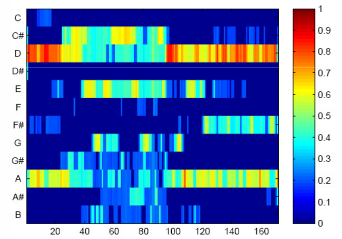
Chroma Features

Example: Bach Toccata

Koopman



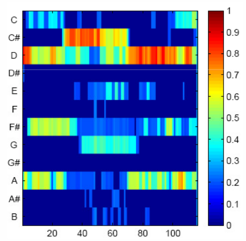
Ruebsam



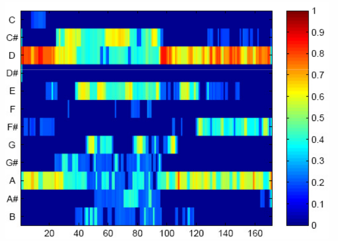
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam

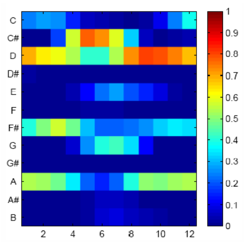


Feature resolution: 10 Hz

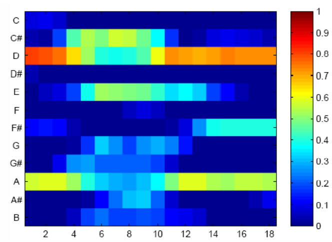
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam

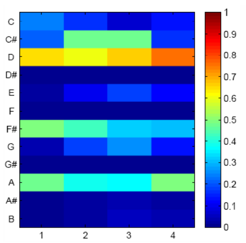


Feature resolution: 1 Hz

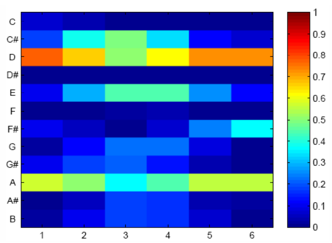
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam



Feature resolution: 0.33 Hz

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
???			
???			
???			

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)	▶	▶	▶
???	▶	▶	▶
???	▶	▶	▶

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)	▶	▶	▶
Beethoven's Fifth (Piano/Sherbakov)	▶	▶	▶
???	▶	▶	▶

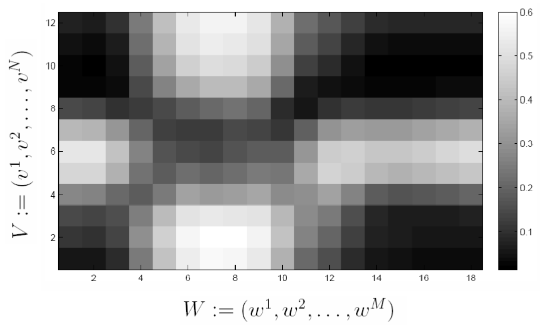
Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)	▶	▶	▶
Beethoven's Fifth (Piano/Sherbakov)	▶	▶	▶
Brahms Hungarian Dance No. 5	▶	▶	▶

Music Synchronization: Audio-Audio

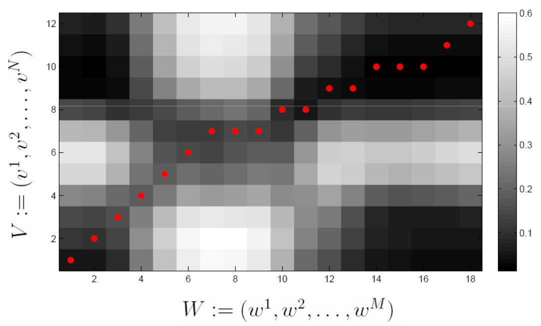
- Koopman $\rightsquigarrow$ $V := (v^1, v^2, \ldots, v^N)$ $N = 12$
  Ruebsam $\rightsquigarrow$ $W := (w^1, w^2, \ldots, w^M)$ $M = 18$
- v^n, w^m = 12-dimensional normalized chroma vectors
- Local cost measure $c : \mathbb{R}^{12} \times \mathbb{R}^{12} \to \mathbb{R}$
  $$c(v^n, w^m) := 1 - \langle v^n, w^m \rangle$$
- $N \times M$ cost matrix $C(n,m) := c(v^n, w^m)$

Music Synchronization: Audio-Audio



Music Synchronization: Audio-Audio

Cost-minimizing warping path



Cost-Minimizing Warping Path

- Computation via dynamic programming

  ⇝ Dynamic Time Warping (DTW)

- Memory requirements and running time: $O(NM)$

- Problem: Infeasible for large N and M

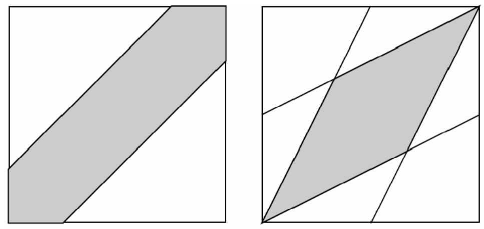
- Example: Feature resolution 10 Hz, pieces 15 min

  $\Rightarrow$ $N, M \sim$ 10,000

  $\Rightarrow$ $N \cdot M \sim$ *100,000,000*

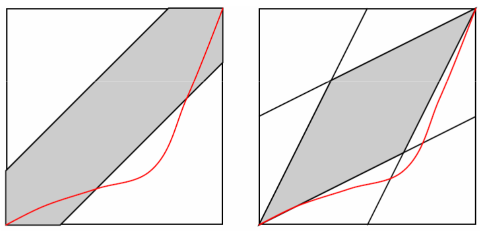
Strategy: Global Constraints

Sakoe-Chiba band   Itakura parallelogram



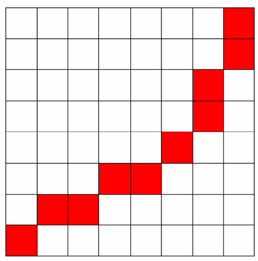
Strategy: Global Constraints

Sakoe-Chiba band   Itakura parallelogram



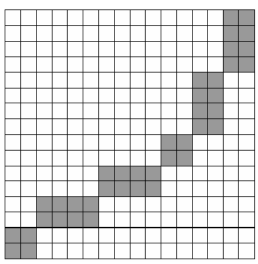
Problem: Optimal warping path not in constraint region

Strategy: Multiscale Approach



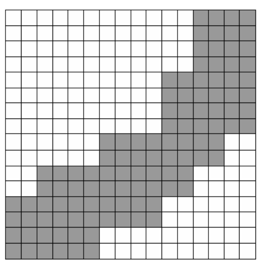
Compute optimal warping path on coarse level

Strategy: Multiscale Approach



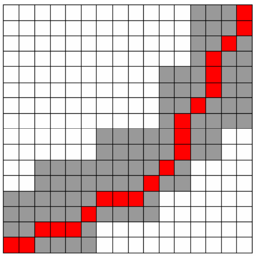
Project on fine level

Strategy: Multiscale Approach



Specify constraint region

Strategy: Multiscale Approach



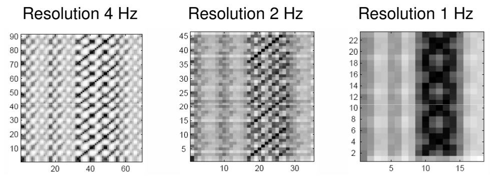
Compute *constrained* optimal warping path

Strategy: Multiscale Approach

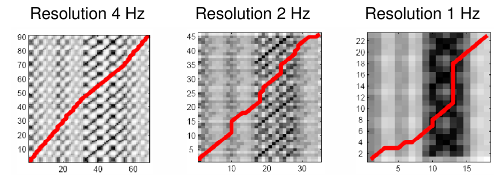
- Suitable features?
- Suitable resolution levels?
- Size of constraint regions?

Good trade-off between efficiency and robustness?

Strategy: Multiscale Approach



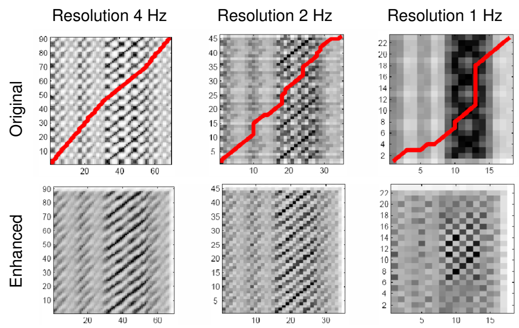
Strategy: Multiscale Approach



Problem: Cost matrix may degenerate
⇝ useless warping path

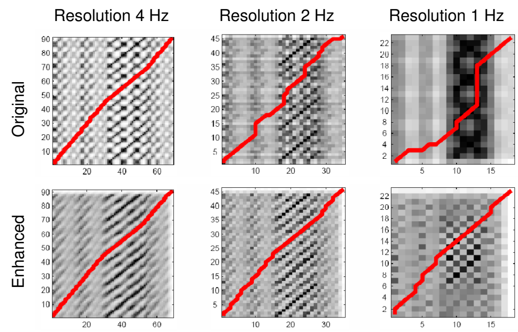
Strategy: Multiscale Approach

Improve robustness by enhancing cost matrix



Strategy: Multiscale Approach

Improve robustness by enhancing cost matrix



Strategy: Multiscale Approach

Chroma features at three levels: 0.33 Hz / 1 Hz / 10 Hz

Recording 1	length [sec]	Recording 2	length [sec]	t_{DTW} [sec]	t_{MsDTW} [sec]	[%]
Beet9Bern	1144.9	Beet9Kar	1054.8	31.18	1.08	3.46

Strategy: Multiscale Approach

Chroma features at three levels: 0.33 Hz / 1 Hz / 10 Hz

Recording 1	length [sec]	Recording 2	length [sec]	t_{DTW} [sec]	t_{MsDTW} [sec]	[%]
Beet9Bern	1144.9	Beet9Kar	1054.8	31.18	1.08	3.46

Number of matrix entries needed for DTW and MsDTW:

	DTW	MsDTW	%
Level 1	120,808,050	2,117,929	1.75
Level 2	1,209,030	17,657	1.46
Level 3	134,464	134,464	100

Music Synchronization: Audio-Audio

Conclusions

- Chroma features

  ⇝ suited for harmony-based music

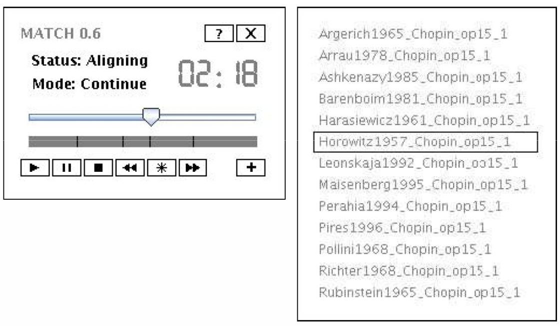
- Relatively coarse but good global alignments
- Multiscale approach: simple, robust, fast

Music Synchronization: Audio-Audio

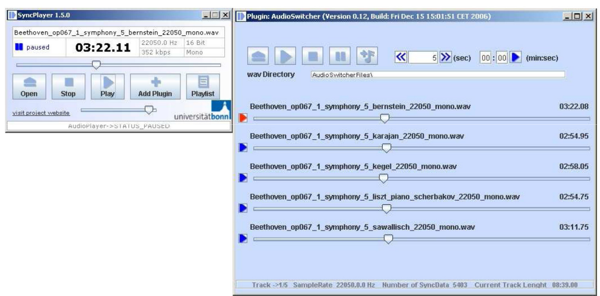
Applications

- Efficient music browsing
- Blending from one interpretation to another one
- Mixing and morphing different interpretations
- Tempo studies

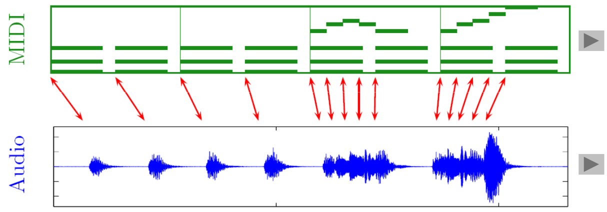
System: Match (Dixon)



System: SyncPlayer/AudioSwitcher



Music Synchronization: MIDI-Audio



Music Synchronization: MIDI-Audio

MIDI = metadata

Automated annotation

Audio recording

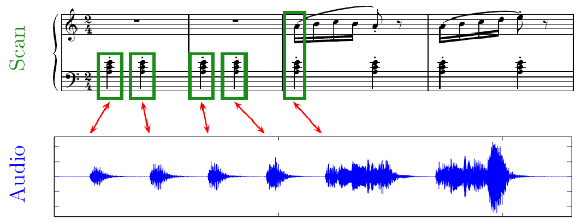
Sonification of annotations

Music Synchronization: MIDI-Audio

Applications

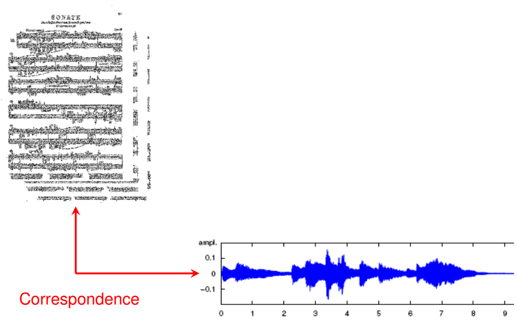
- Automated audio annotation
- Accurate audio access after MIDI-based retrieval
- Automated tracking of MIDI note parameters during audio playback

Music Synchronization: Scan-Audio



Music Synchronization: Scan-Audio

Scanned Sheet Music

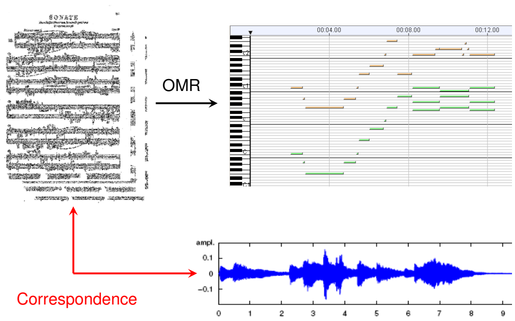


Audio Recording

Music Synchronization: Scan-Audio

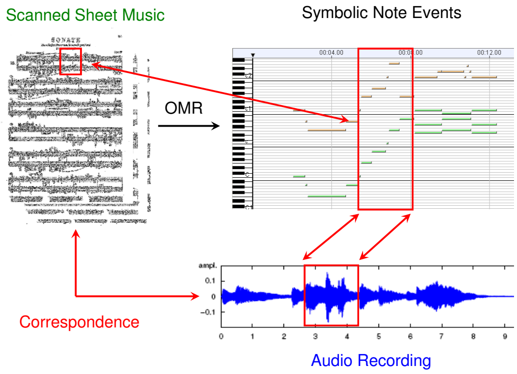
Scanned Sheet Music

Symbolic Note Events

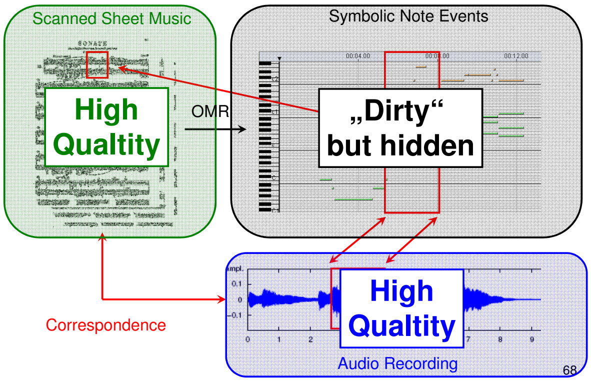


Audio Recording

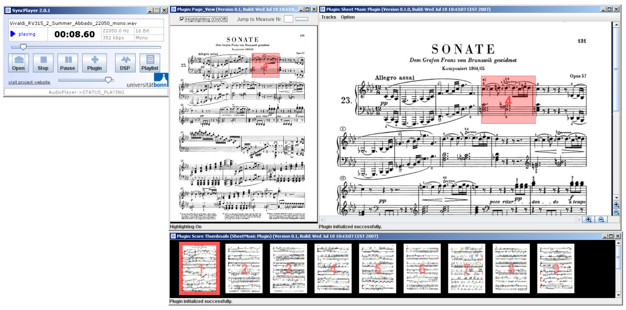
Music Synchronization: Scan-Audio



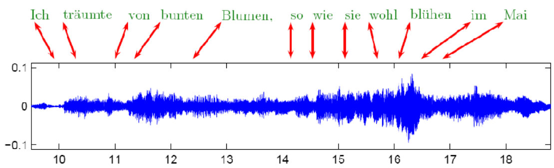
Music Synchronization: Scan-Audio



System: SyncPlayer/SheetMusic



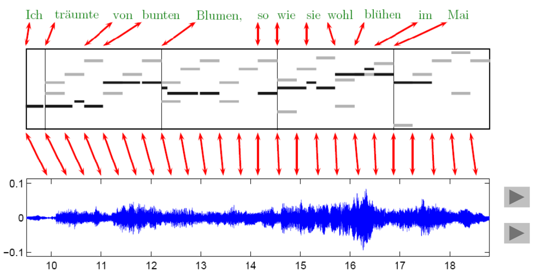
Music Synchronization: Lyrics-Audio



Difficult task!

Music Synchronization: Lyrics-Audio

Lyrics-Audio → Lyrics-MIDI + MIDI-Audio



System: SyncPlayer/LyricsSeeker

Conclusions: Music Synchronization

Various requirements

- Efficiency
- Robustness
- Accuracy
- Variablity of music

Conclusions: Music Synchronization

Combination of various strategies

- Feature level
- Local cost measure level
- Global alignment level
- Evidence pooling using competing strategies

Conclusions: Music Synchronization

Combination of various strategies

- Feature level
- Local cost measure level
- Global alignment level
- Evidence pooling using competing strategies

Example: MIDI-Audio synchronization

Chroma-Chroma:
Chroma-Chroma + onset-bonus:

Conclusions: Music Synchronization

Offline vs. Online

- Online version: Dixon/Widmer (ISMIR 2005)
- Hidden Markov Models: Raphael (ISMIR 2004)
- Score-following
- Automatic accompaniment

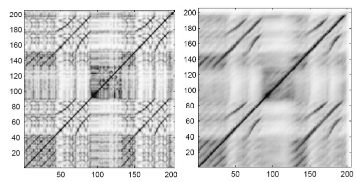
Conclusions: Music Synchronization

Presence of variations

- Instrumentation
- Musical structure
- Polyphony
- Musical key
- …

Part II

Audio Structure Analysis



Music Structure Analysis

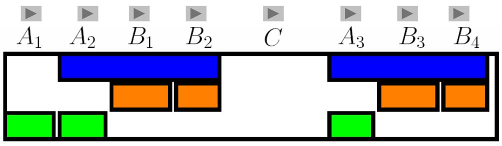
- Music segmentation
  - pitch content (e.g., melody, harmony)
  - music texture (e.g., timbre, instrumentation, sound)
  - rhythm
- Detection of repeating sections, phrases, motives
  - song structure (e.g., intro, versus, chorus)
  - musical form (e.g., sonata, symphony, concerto)
- Detection of other hidden relationships

Audio Structure Analysis

Given: CD recording

Goal: Automatic extraction of the repetitive structure
(or of the musical form)

Example: Brahms Hungarian Dance No. 5 (Ormandy)



Audio Structure Analysis

- Dannenberg/Hu (ISMIR 2002)
- Peeters/Burthe/Rodet (ISMIR 2002)
- Cooper/Foote (ISMIR 2002)
- Goto (ICASSP 2003)
- Chai/Vercoe (ACM Multimedia 2003)
- Lu/Wang/Zhang (ACM Multimedia 2004)
- Bartsch/Wakefield (IEEE Trans. Multimedia 2005)
- Goto (IEEE Trans. Audio 2006)
- Müller/Kurth (EURASIP 2007)
- Rhodes/Casey (ISMIR 2007)
- Peeters (ISMIR 2007)

Audio Structure Analysis

- Audio features
- Cost measure and cost matrix

  $\rightsquigarrow$ self-similarity matrix
- Path extraction (pairwise similarity of segments)
- Global structure (clustering, grouping)

Audio Structure Analysis

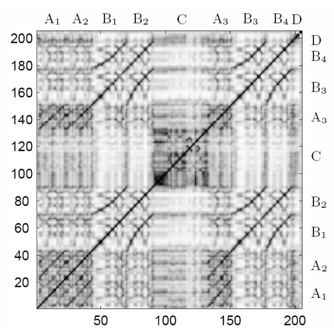
- Audio $\rightsquigarrow$ $V := (v^1, v^2, \ldots, v^N)$
- v^n = 12-dimensional normalized chroma vector
- Local cost measure $c : \mathbb{R}^{12} \times \mathbb{R}^{12} \to \mathbb{R}$

  $$c(v^n, w^m) := 1 - \langle v^n, w^m \rangle$$
- $N \times N$ cost matrix $C(n,m) := c(v^n, w^m)$

  $\rightsquigarrow$ quadratic self-similarity matrix

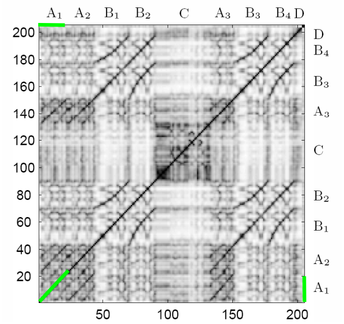
Audio Structure Analysis

Self-similarity matrix



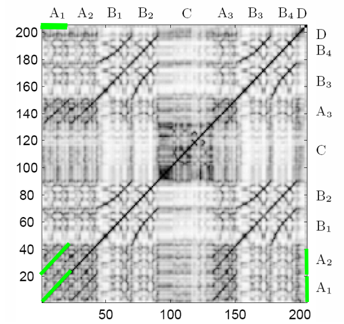
Audio Structure Analysis

Self-similarity matrix



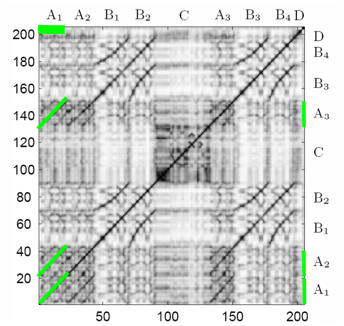
Audio Structure Analysis

Self-similarity matrix



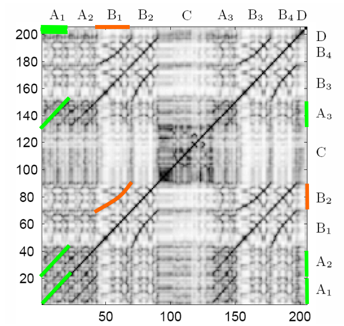
Audio Structure Analysis

Self-similarity matrix



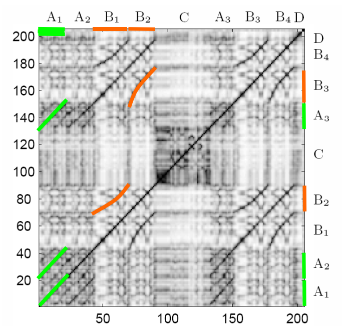
Audio Structure Analysis

Self-similarity matrix



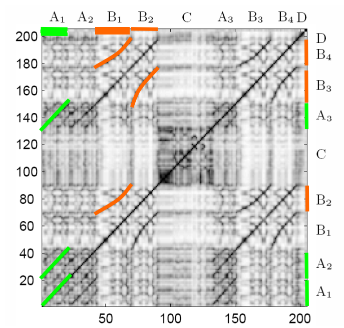
Audio Structure Analysis

Self-similarity matrix



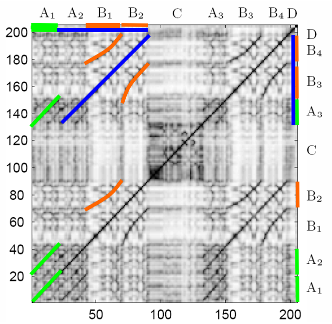
Audio Structure Analysis

Self-similarity matrix



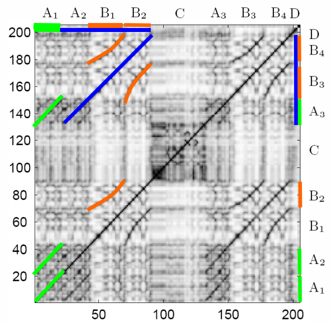
Audio Structure Analysis

Self-similarity matrix

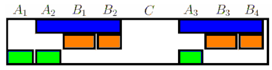


Audio Structure Analysis

Self-similarity matrix



Similarity cluster



Matrix Enhancement

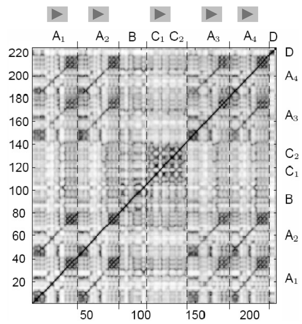
Challenge: Presence of musical variations

- Fragmented paths and gaps
- Paths of poor quality
- Regions of constant (low) cost
- Curved paths

Idea: Enhancement of path structure

Matrix Enhancement

Shostakovich Waltz 2, Jazz Suite No. 2 (Chailly)



Matrix Enhancement

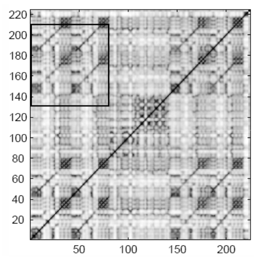
Idea: Usage of contextual information (Foote 1999)

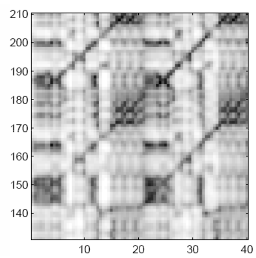
$$C_L(n,m) := \frac{1}{L}\sum_{\ell=0}^{L-1} c(v_{n+\ell}, v_{m+\ell})$$

- Comparison of entire sequences
- $L =$ length of sequences
- $C_L =$ enhanced cost matrix

⇝ smoothing effect

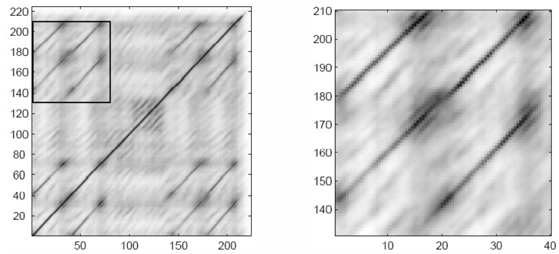
Matrix Enhancement (Shostakovich)





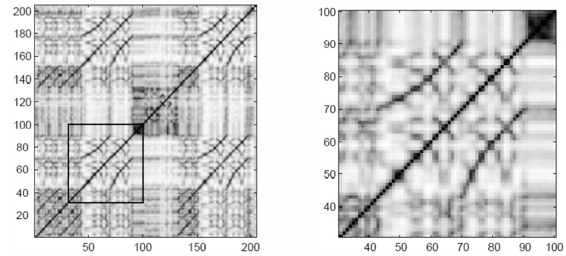
Cost matrix C

Matrix Enhancement (Shostakovich)



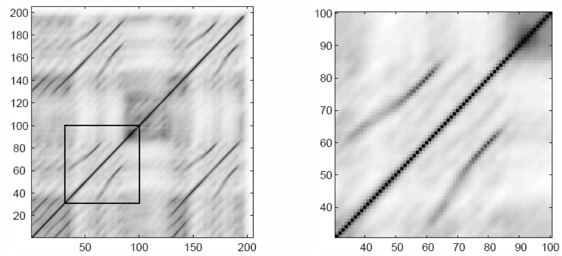
Enhanced cost matrix C_L

Matrix Enhancement (Brahms)



Cost matrix C

Matrix Enhancement (Brahms)



Enhanced cost matrix C_L

Problem: Relative tempo differences are smoothed out

Matrix Enhancement

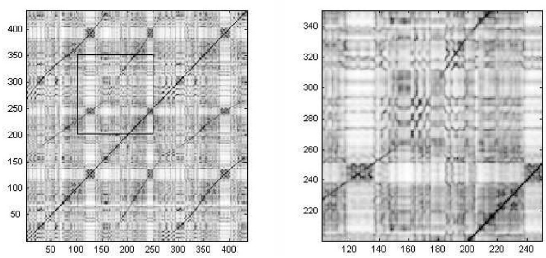
Idea: Smoothing along various directions
and minimizing over all directions

$$C_L^{\min}(n,m) := \min_k C_L^{\mathrm{slope}_k}(n,m)$$

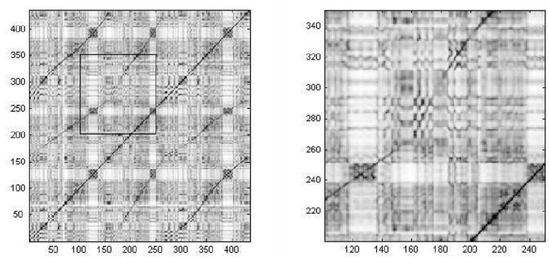
- $\mathrm{slope}_k =$ kth direction of smoothing
- $C_L^{\mathrm{slope}_k} =$ enhanced cost matrix w.r.t. slope_k
- Usage of eight slope values

$\rightsquigarrow$ tempo changes of -30 to +40 percent

Matrix Enhancement

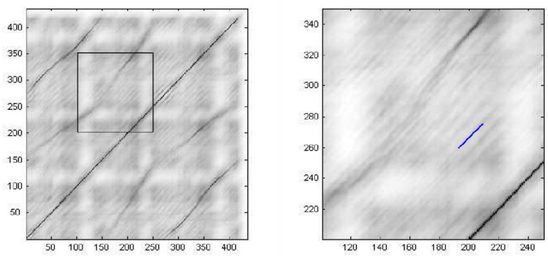


Matrix Enhancement



Cost matrix C

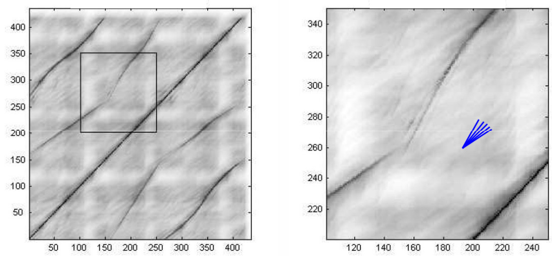
Matrix Enhancement



Cost matrix C_L with $L = 20$

Filtering along main diagonal

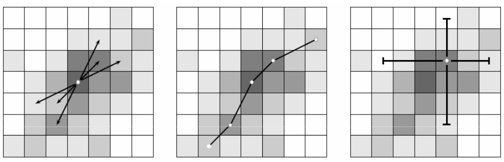
Matrix Enhancement



Cost matrix $C_L^{\min}$ with $L = 20$

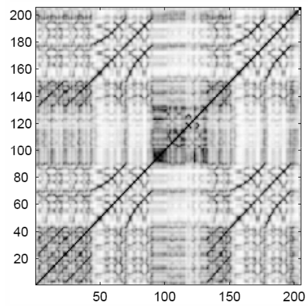
Filtering along 8 different directions and minimizing

Path Extraction



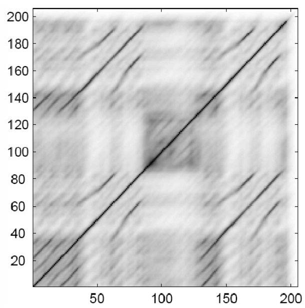
- Start with initial point
- Extend path in greedy fashion
- Remove path neighborhood

Path Extraction



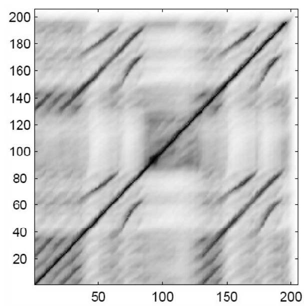
Cost matrix C

Path Extraction



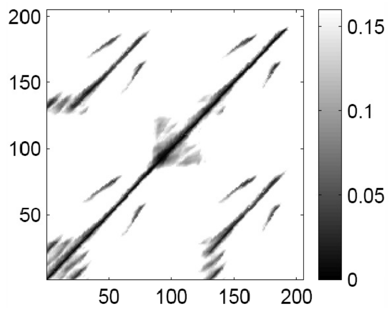
Enhanced cost matrix C_L

Path Extraction



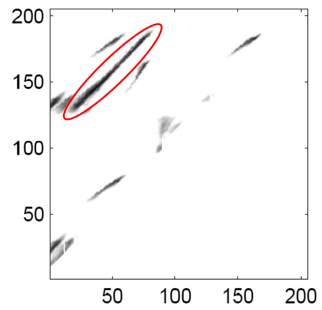
Enhanced cost matrix $C_L^{\min}$

Path Extraction



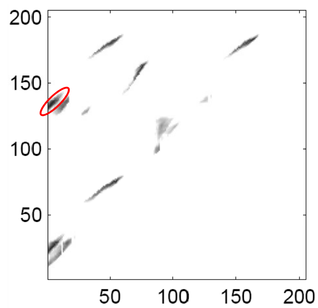
Thresholded $C_L^{\min}$

Path Extraction



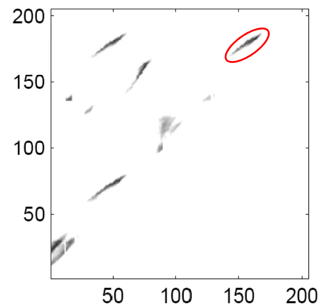
Thresholded $C_L^{\min}$, upper left

Path Extraction



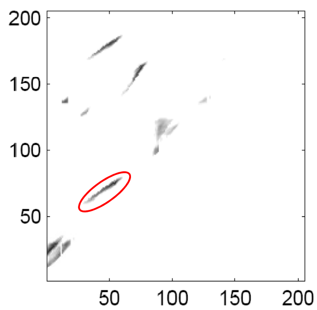
Path removal

Path Extraction



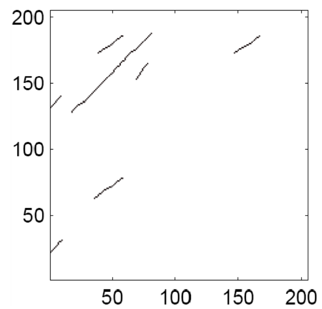
Path removal

Path Extraction



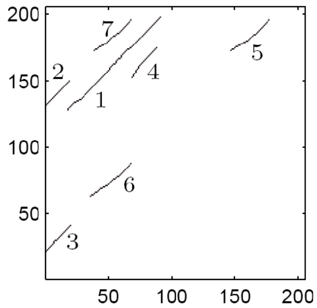
Path removal

Path Extraction



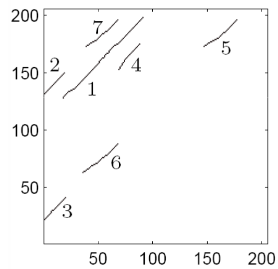
Extracted paths

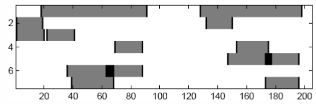
Path Extraction



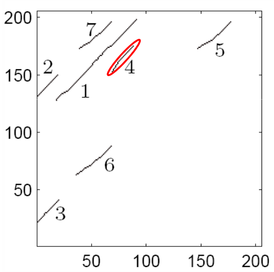
Extracted paths after postprocessing

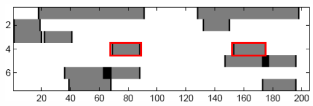
Global Structure





Global Structure





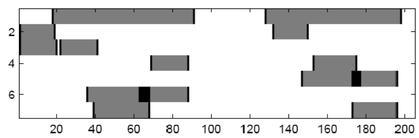
How can one derive the global structure from pairwise relations?

Global Structure

- Taks: Computation of similarity clusters
- Problem: Missing and inconsistent path relations
- Strategy: Approximate “transitive hull”

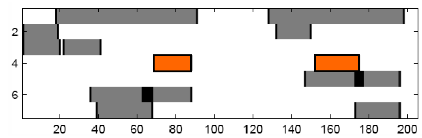
Global Structure

Path relations



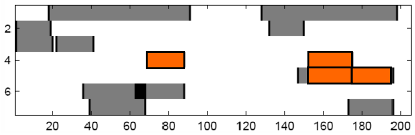
Global Structure

Path relations



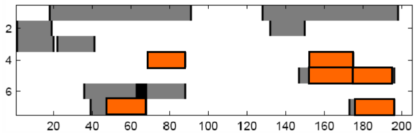
Global Structure

Path relations



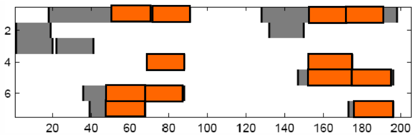
Global Structure

Path relations



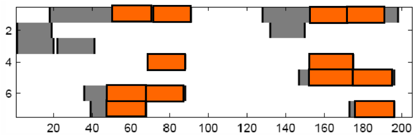
Global Structure

Path relations

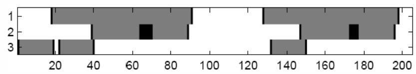


Global Structure

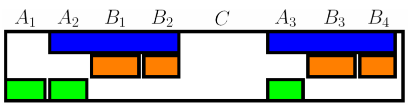
Path relations



Final result

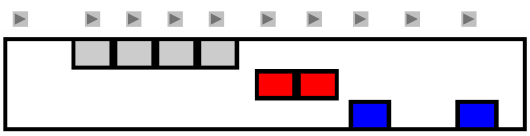


Ground truth



Transposition Invariance

Example: Zager & Evans “In The Year 2525”



Transposition Invariance

Goto (ICASSP 2003)

- Cyclically shift chroma vectors in one sequence
- Compare shifted sequence with original sequence
- Perform for each of the twelve shifts a separate structure analysis
- Combine the results

Transposition Invariance

Goto (ICASSP 2003)

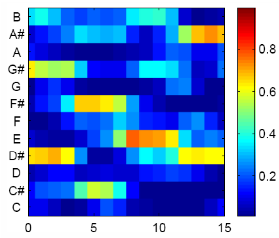
- Cyclically shift chroma vectors in one sequence
- Compare shifted sequence with original sequence
- Perform for each of the twelve shifts a separate structure analysis
- Combine the results

Müller/Clausen (ISMIR 2007)

- Integrate all cyclic information in one transposition-invariant self-similarity matrix
- Perform one joint structure analysis

Transposition Invariance

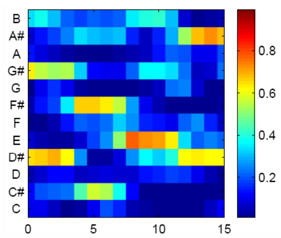
Example: Zager & Evans "In The Year 2525"



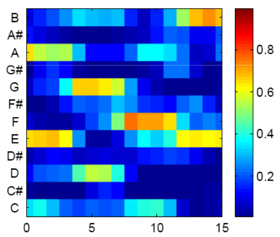
Original: $(v^1, \ldots, v^N)$

Transposition Invariance

Example: Zager & Evans "In The Year 2525"

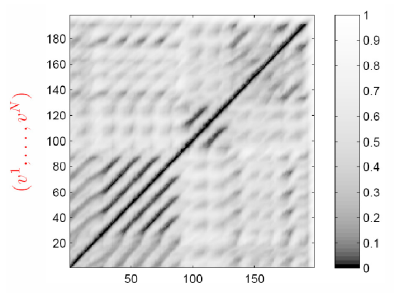


Original: $(v^1, \ldots, v^N)$

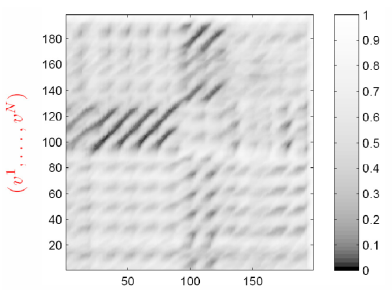


Shifted: $(\sigma(v^1), \ldots, \sigma(v^N))$

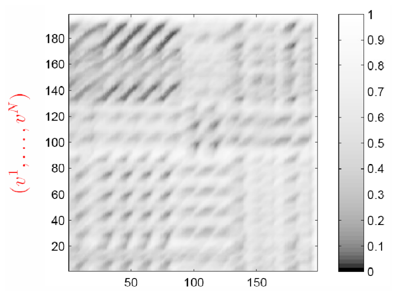
Transposition Invariance



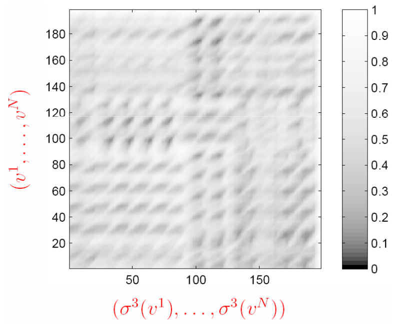
Transposition Invariance



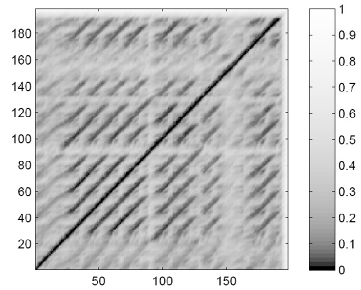
Transposition Invariance



Transposition Invariance

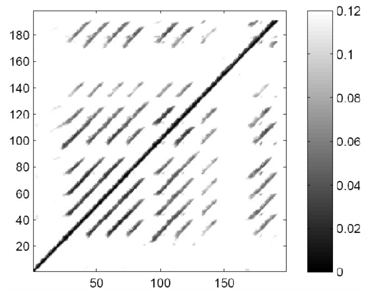


Transposition Invariance



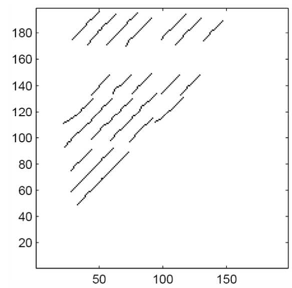
Minimize over all twelve matrices

Transposition Invariance



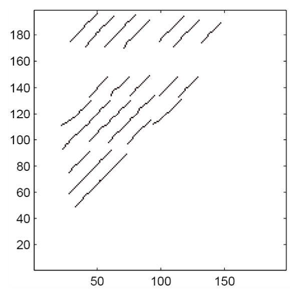
Thresholded self-similarity matrix

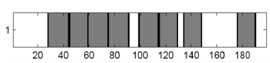
Transposition Invariance



Path extraction

Transposition Invariance



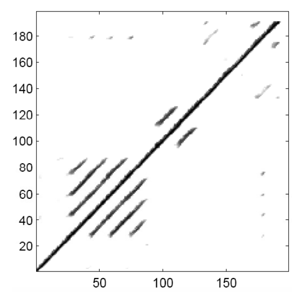


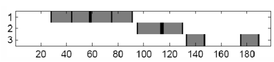
Path extraction

Computation of similarity clusters

Transposition Invariance

Stabilizing effect

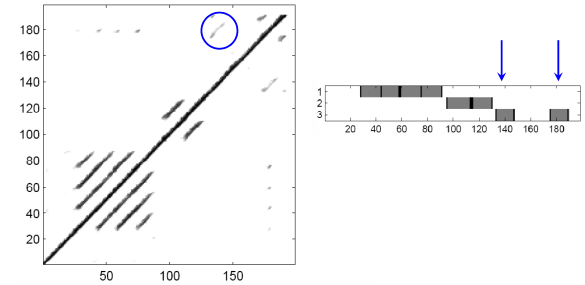




Self-similarity matrix (thresholded)

Transposition Invariance

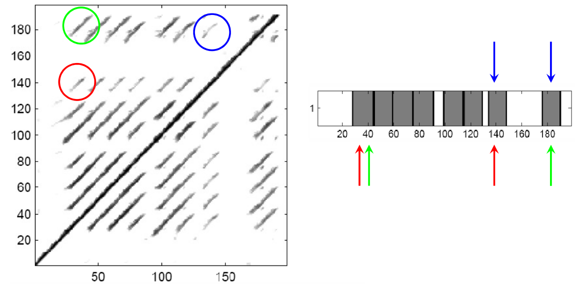
Stabilizing effect



Self-similarity matrix (thresholded)

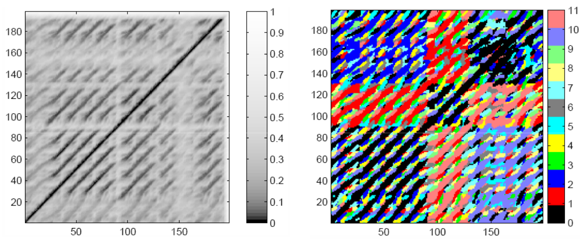
Transposition Invariance

Stabilizing effect



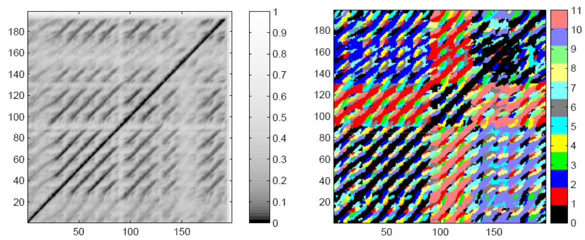
Transposition-invariant self-similarity matrix (thresholded)

Transposition Invariance



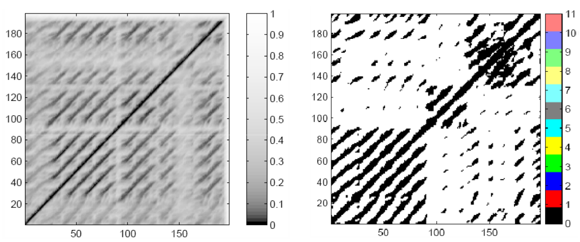
Transposition-invariant matrix   Minimizing shift index

Transposition Invariance



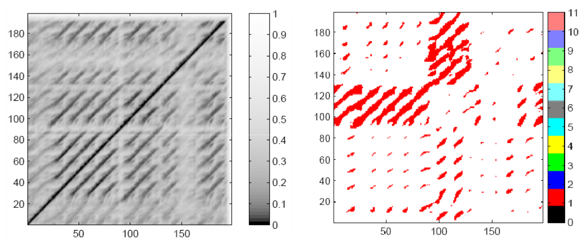
Transposition-invariant matrix   Minimizing shift index

Transposition Invariance



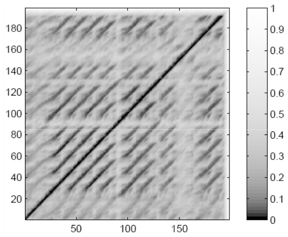
Transposition-invariant matrix   Minimizing shift index = 0

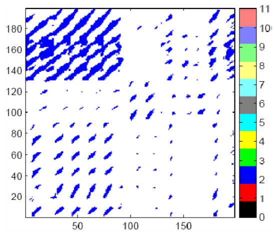
Transposition Invariance



Transposition-invariant matrix   Minimizing shift index = 1

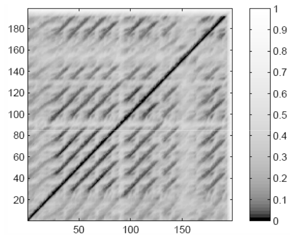
Transposition Invariance

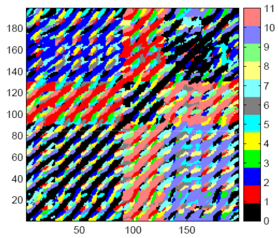




Transposition-invariant matrix   Minimizing shift index = 2

Transposition Invariance



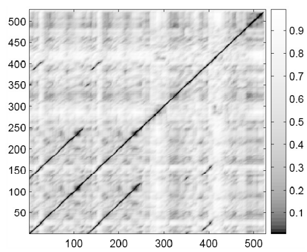


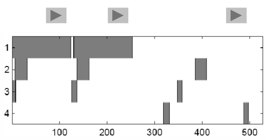
Serra/Gomez (ICASSP 2008): Used for Cover Song ID

Discrete structure ⇝ suitable for indexing?

Transposition Invariance

Example: Beethoven “Tempest”

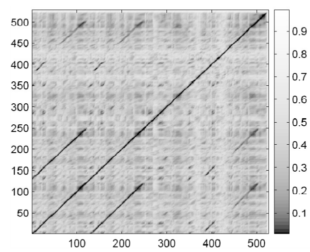


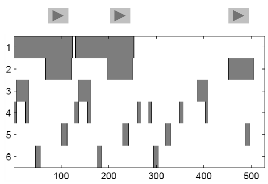


Self-similarity matrix

Transposition Invariance

Example: Beethoven “Tempest”





Transposition-invariant self-similarity matrix

Conclusions: Audio Structure Analysis

Challenge: Musical variations

- Timbre, dynamics, tempo
- Musical key ⇝ cyclic chroma shifts
- Major/minor
- Differences at note level / improvisations

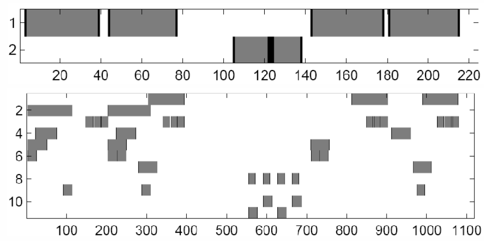
Conclusions: Audio Structure Analysis

Strategy: Matrix enhancement

- Filtering techniques / contextual information
  - Cooper/Foote (ISMIR 2002)
  - Müller/Kurth (ICASSP 2006)
- Transposition-invariant similarity matrices
  - Goto (ICASSP 2003)
  - Müller/Clausen (ISMIR 2007)
- Higher-order similarity matrices
  - Peeters (ISMIR 2007)

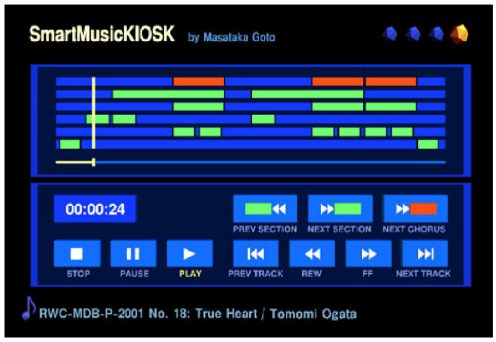
Conclusions: Audio Structure Analysis

Challenge: Hierarchical structure of music

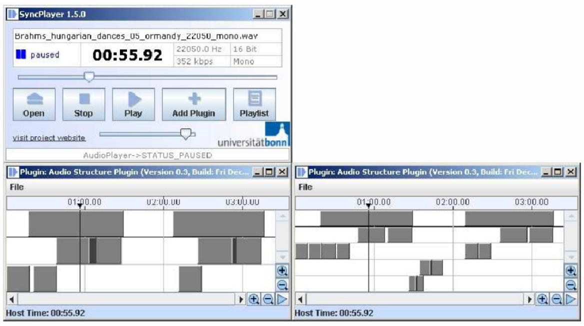


Rhodes/Casey (ISMIR 2007)

System: SmartMusicKiosk (Goto)

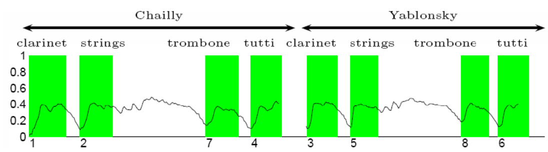


System: SyncPlayer/AudioStructure



Part III

Audio Matching



Audio Matching

Given: Large music database containing several
- recordings of the same piece of music
- interpretations by various musicians
- arrangements in different instrumentations

Goal: Given a short query audio clip, identify all corresponding audio clips of similar musical content
- irrespective of the specific interpretation and instrumentation
- automatically and efficiently

Query-by-Example paradigm

Audio Matching

- Müller/Kurth/Clausen (ISMIR 2005)
- Kurth/Müller (IEEE T-ASLP 2008)

Related problems

Audio identification

- Allamanche et al. (AES 2001)
- Cano et al. (*IEEE MMSP* 2002)
- Kurth/Clausen/Ribbrock (AES 2002)
- Wang (ISMIR 2003)
- Shrestha/Kalker (ISMIR 2004)

Audio synchronization

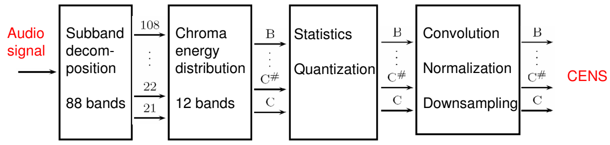
Audio structure analysis

Audio Matching

General strategy

- Normalized and smoothed chroma features
  - correlates to harmonic progression
  - robust to variations in dynamics, timbre, articulation, local tempo
- Robust matching procedure
  - efficient
  - robust to global tempo variations
  - scalable using index structure

Feature Design



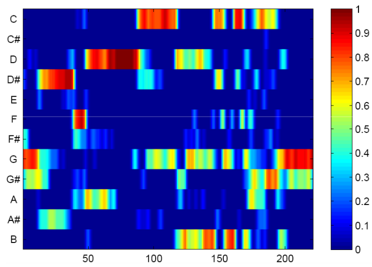
Two stages:

Stage 1: Local chroma energy distribution features
Stage 2: Normalized short-time statistics

⇝ CENS = Chroma Energy Normalized Statistics

Feature Design

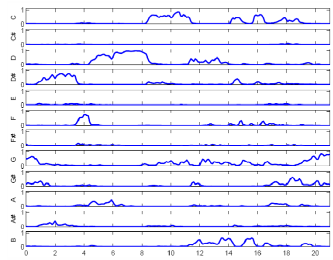
Beethoven‘s Fifth: Bernstein



Resolution: 10 features/second
Feature window size: 200 milliseconds

Feature Design

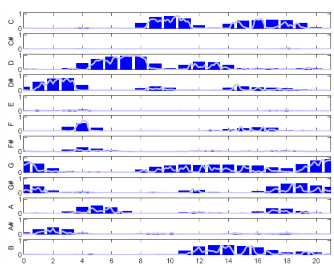
Beethoven‘s Fifth: Bernstein



Resolution: 10 features/second
Feature window size: 200 milliseconds

Feature Design

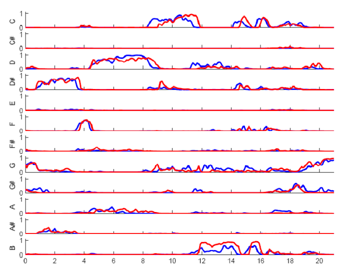
Beethoven‘s Fifth: Bernstein



Resolution: 1 features/second
Feature window size: 4000 milliseconds

Feature Design

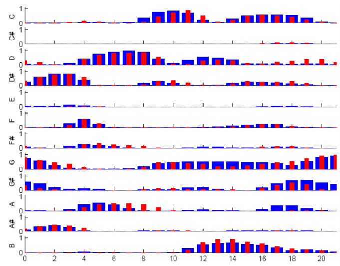
Beethoven‘s Fifth: Bernstein vs. Sawallisch



Resolution: 10 features/second
Feature window size: 200 milliseconds

Feature Design

Beethoven's Fifth: Bernstein vs. Sawallisch

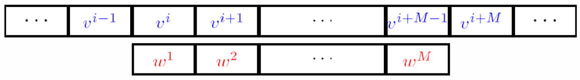


Resolution: 1 features/second
Feature window size: 4000 milliseconds

Matching Procedure

Compute CENS feature sequences

- Database $D \rightsquigarrow F[D] = (v^1, v^2, \dots, v^N)$
- Query $Q \rightsquigarrow F[Q] = (w^1, w^2, \dots, w^M)$
- $N \approx 500000,\ M \approx 20$

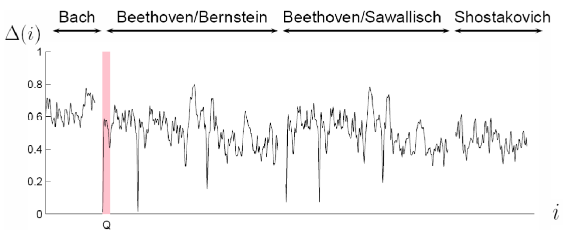


$\Delta(i) :=$ local distance$((v^i, v^{i+1}, \dots, v^{i+M-1}), (w^1, w^2, \dots, w^M))$

$\rightsquigarrow$ Global distance function $\Delta : [1:N] \rightarrow [0,1]$

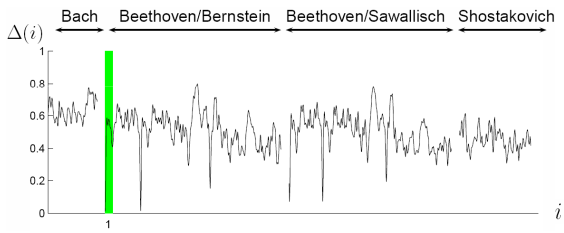
Matching Procedure

Query: Beethoven's Fifth / Bernstein, first 20 seconds



Matching Procedure

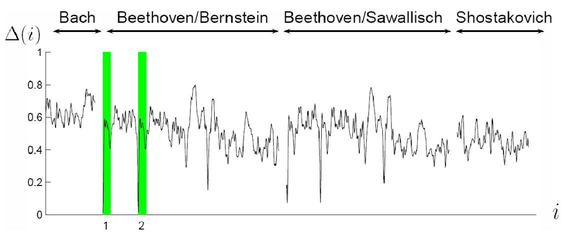
Query: Beethoven's Fifth / Bernstein, first 20 seconds



Best audio matches: 1

Matching Procedure

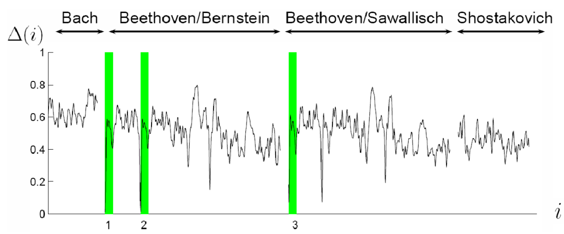
Query: Beethoven's Fifth / Bernstein, first 20 seconds



Best audio matches: 2

Matching Procedure

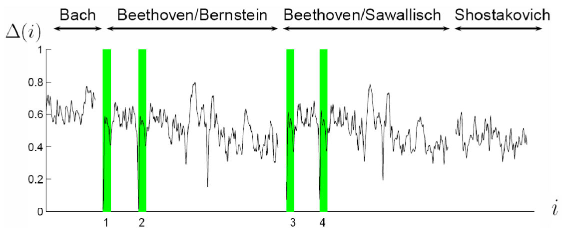
Query: Beethoven's Fifth / Bernstein, first 20 seconds



Best audio matches: 3

Matching Procedure

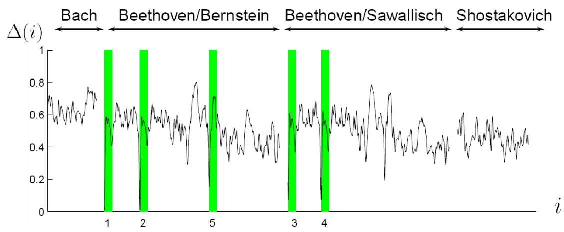
Query: Beethoven‘s Fifth / Bernstein, first 20 seconds



Best audio matches: 4

Matching Procedure

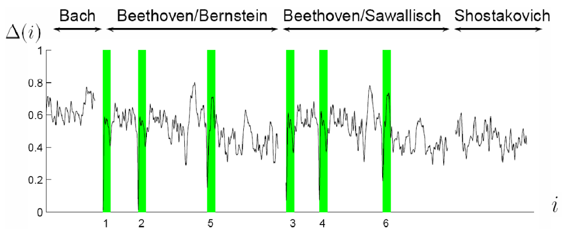
Query: Beethoven‘s Fifth / Bernstein, first 20 seconds



Best audio matches: 5

Matching Procedure

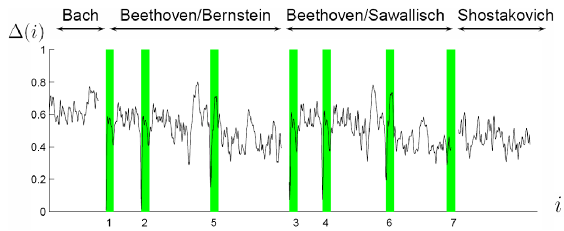
Query: Beethoven‘s Fifth / Bernstein, first 20 seconds



Best audio matches: 6

Matching Procedure

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds



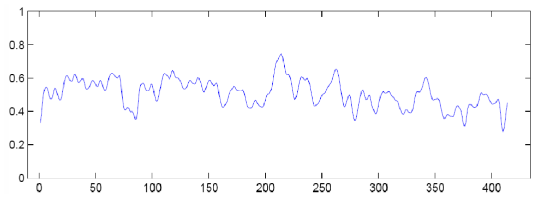
Best audio matches: 7

Global Tempo Variations

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds

Problem: Karajan is much faster $\rightsquigarrow$ useless Δ

Solution?

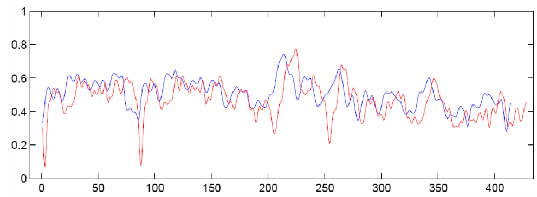


Global Tempo Variations

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds

Problem: Karajan is much faster $\rightsquigarrow$ useless Δ

Solution: Make Bernstein query faster and comute new Δ

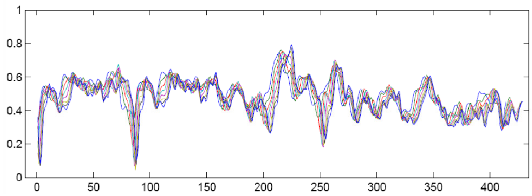


Global Tempo Variations

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds

Problem: Karajan is much faster $\rightsquigarrow$ useless Δ

Solution: Compute Δ for various tempi

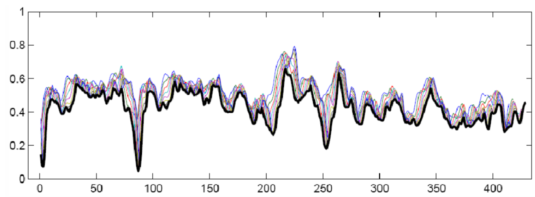


Global Tempo Variations

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds

Problem: Karajan is much faster $\rightsquigarrow$ useless Δ

Solution: Minimize over all resulting Δ's $\rightsquigarrow$ $\Delta^{\min}$

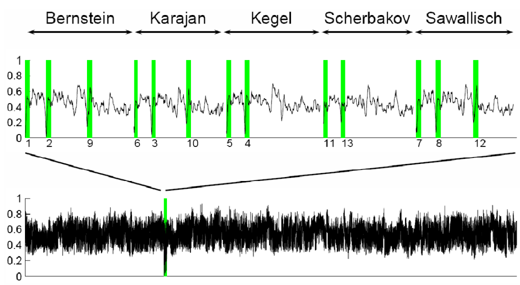


Experiments

- Audio database > 110 hours, 16.5 GB
- Preprocessing $\rightsquigarrow$ CENS features, 40.3 MB
- Query clip $\approx$ 20 seconds
- Query response time < 10 seconds

Experiments

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds



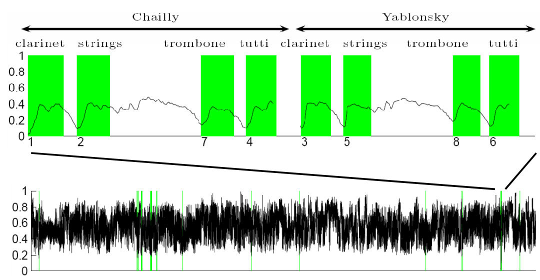
Experiments

Query: Beethoven‘s Fifth / Bernstein, first 20 seconds

Rank	$\Delta^{\min}$	Piece	Position
1	0.0114	Beethoven's Fifth/Bernstein	0 - 21
2	0.0150	Beethoven's Fifth/Bernstein	101 - 122
3	0.0438	Beethoven's Fifth/Karajan	86 - 103
⋮	⋮	⋮	⋮
10	0.1796	Beethoven's Fifth/Karajan	252 - 271
11	0.1827	Beethoven (Liszt) Fifth/Scherbakov	0 - 19
12	0.1945	Beethoven's Fifth/Sawallisch	275 - 296
13	0.1970	Beethoven's Fifth (Liszt)/Scherbakov	86 - 103
14	0.2169	Schumann op 97,1/Levine	28 - 43
⋮	⋮	⋮	⋮

Experiments

Query: Shostakovich, Waltz/Chailly, first 27 seconds



Experiments

Query: Shostakovich, Waltz/Chailly, first 21 seconds

Rank	$\Delta^{\min}$	Piece	Position
1	0.0172	Shostakovich/Chailly	0 - 21
2	0.0505	Shostakovich/Chailly	41 - 60
3	0.0983	Shostakovich/Chailly	180 - 198
4	0.1044	Shostakovich/Yablonsky	1 - 19
5	0.1090	Shostakovich/Yablonsky	36 - 52
6	0.1401	Shostakovich/Yablonsky	156 - 174
7	0.1476	Shostakovich/Chailly	144 - 162
8	0.1626	Bach BWV 582/Chorzempa	358 - 373
9	0.1668	Beethoven op 37,1/Toscanini	12 - 28
10	0.1729	Beethoven op 37,1/Pollini	202 - 218
⋮	⋮	⋮	⋮

Conclusions

Strategy: Absorb variations at feature level

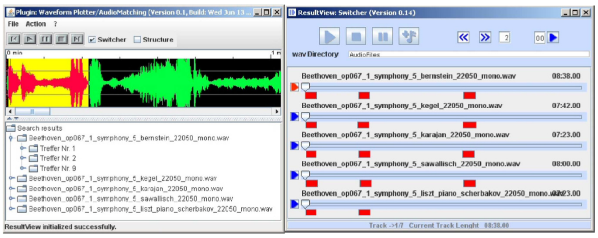
- Chroma ⇝ invariance to timbre
- Normalization ⇝ invariance to dynamics
- Smoothing ⇝ invariance to local time deviations

Conclusions

Global Matching Procedure

- Strategy: Exact matching and multiple scaled queries
  - simulate tempo variations by feature resampling
  - different queries correspond to different tempi
  - indexing possible
- Strategy: Dynamic Time Warping
  - subsequence variant
  - more flexible (in particular for longer queries)
  - indexing hard

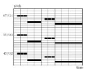
System: SyncPlayer/AudioMatching



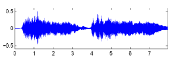
Multimodal Computing and Interaction

Sheet Music (Image)

MIDI



CD / MP3 (Audio)



MusicXML (Text)

Music

Singing / Voice (Audio)

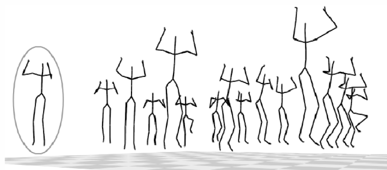
Music Literature (Text)

Music Film (Video)

Dance / Motion (Mocap)

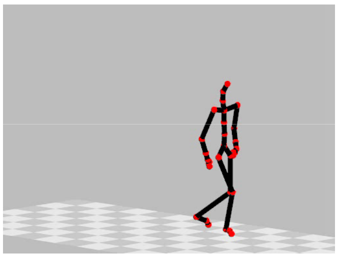
Part IV

Motion Retrieval



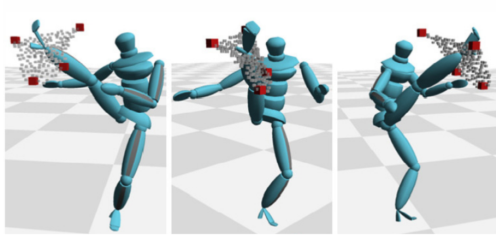
Motion Capture Data

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



Motion Capture Data

Application: Motion Morphing



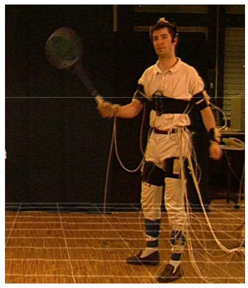
From Kovar/Gleicher (SIGGRAPH 2004)

Motion Capture Data

Optical System

Motion Capture Data

Mechanical and magnetic systems

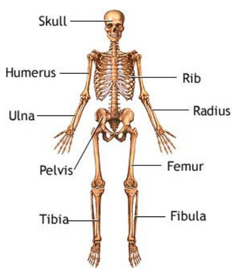


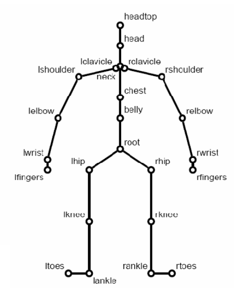
http://www.metamotion.com/gypsy/gypsy-motion-capture-system.htm

http://vrlab.epfl.ch/research/MC_motion_capture.html

Motion Capture Data

Skeletal kinematic chain

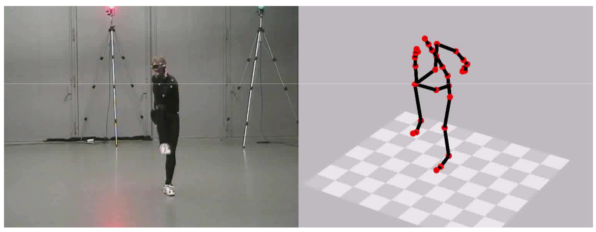




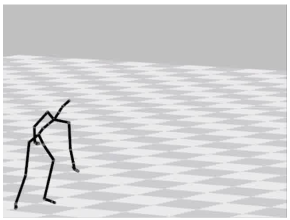
http://apps.uwhealth.org/health/adam/graphics/images/en/9065.jpg

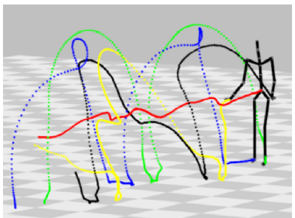
Motion Capture Data

Conversion: Marker → Skeleton



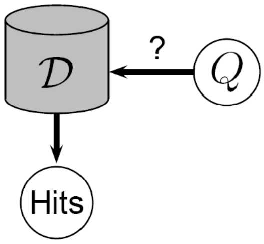
Motion Capture Data



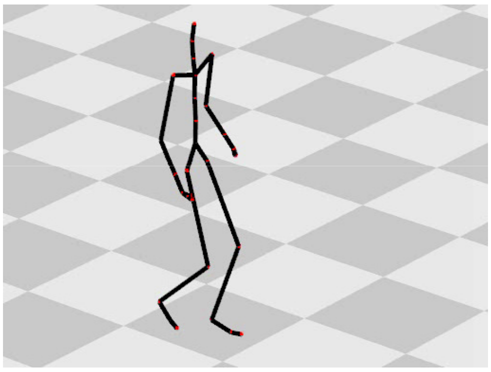


Motion Retrieval

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

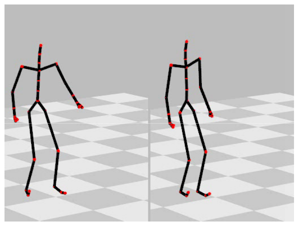


Motion Retrieval



Motion Similarity

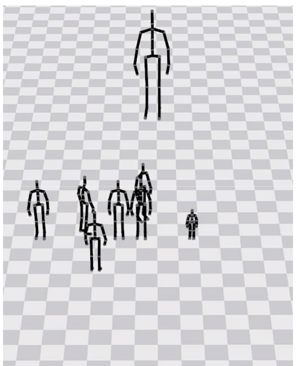
- Numerical similarity vs. logical similarity
- Logically related motions may exhibit significant spatio-temporal variations



Motion Similarity

Global Transforms

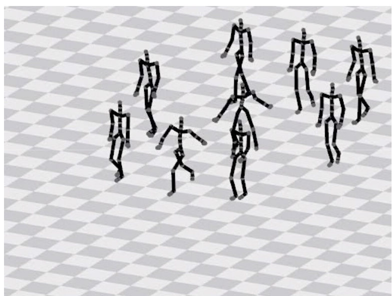
- Translation
- Spatial scaling
- Rotation
- Reflection
- Temporal Scaling



Motion Similarity

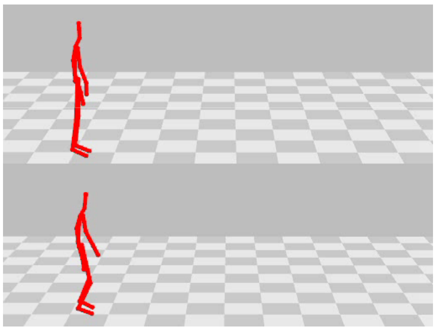
Motion Styles

- Cheerful walking
- Furious walking
- Limping
- Tiptoeing
- Marching



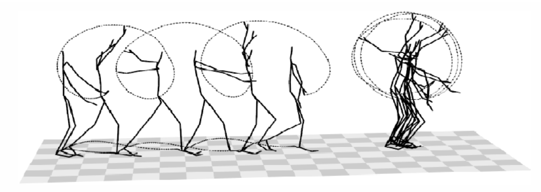
Motion Similarity

Spatio-Temporal Deformations



Motion Similarity

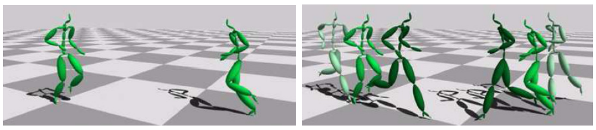
Partial Similarity



Local Similarity Measure

Point cloud (Kovar & Gleicher)

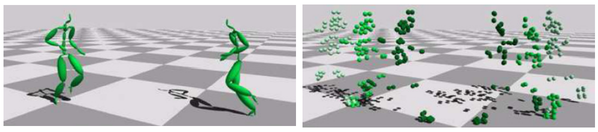
$$c^{3D}\Big(D(n), D(m)\Big) := \min_{\theta,x,z} \left(\sum_{i=1}^{K} w_i \|p_i - T_{\theta,x,z}(p_i')\|^2 \right)$$



Local Similarity Measure

Point cloud (Kovar & Gleicher)

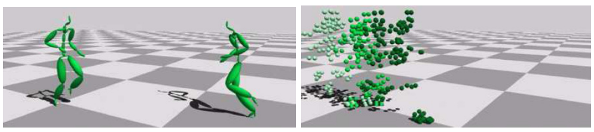
$$c^{3D}\Big(D(n), D(m)\Big) := \min_{\theta,x,z} \left(\sum_{i=1}^{K} w_i \|p_i - T_{\theta,x,z}(p_i')\|^2 \right)$$



Local Similarity Measure

Point cloud (Kovar & Gleicher)

$$c^{3D}\Big(D(n), D(m)\Big) := \min_{\theta,x,z} \left(\sum_{i=1}^{K} w_i \|p_i - T_{\theta,x,z}(p_i')\|^2 \right)$$



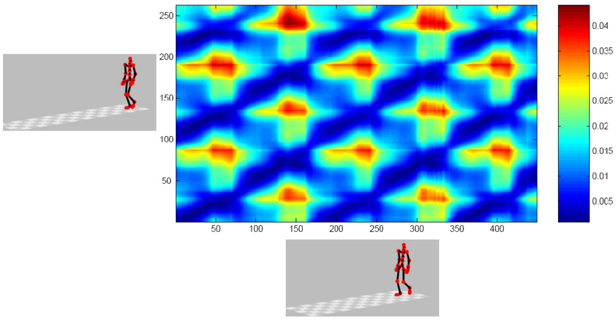
Local Similarity Measure

Quaternions

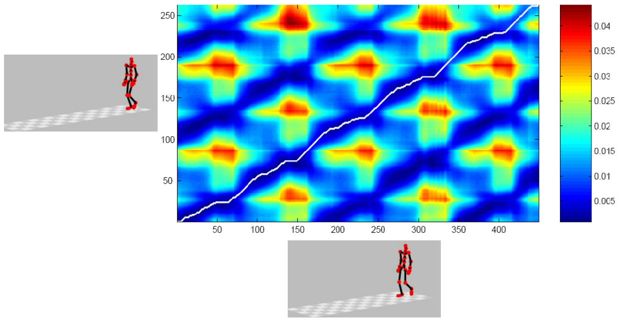
$$c^{\mathrm{Quat}} : \mathcal{J} \times \mathcal{J} \to [0, 1]$$

$$c^{\mathrm{Quat}}(j, j') := \sum_{b \in B} w_b \cdot \frac{2}{\pi} \cdot \arccos |\langle q_b | q_b' \rangle|$$

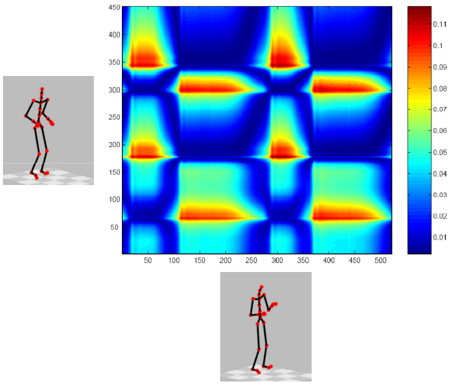
Dynamic Time Warping (DTW)



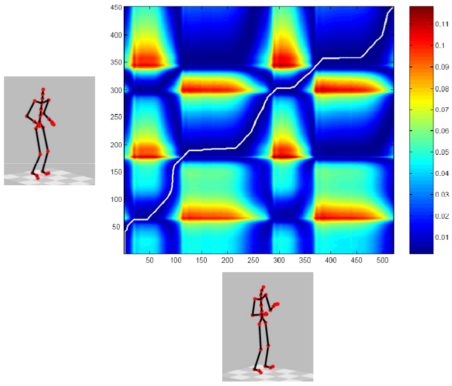
Dynamic Time Warping (DTW)



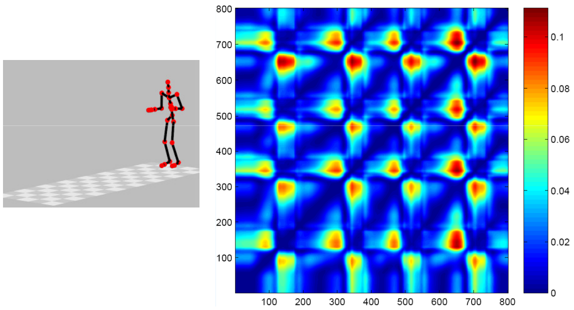
Dynamic Time Warping (DTW)



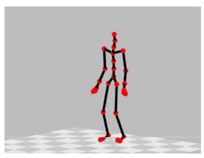
Dynamic Time Warping (DTW)

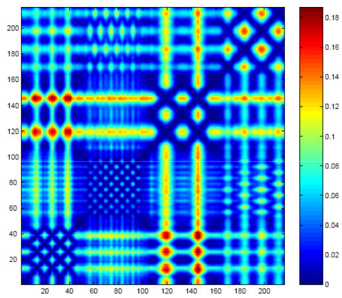


Self-Similarity Matrix



Self-Similarity Matrix

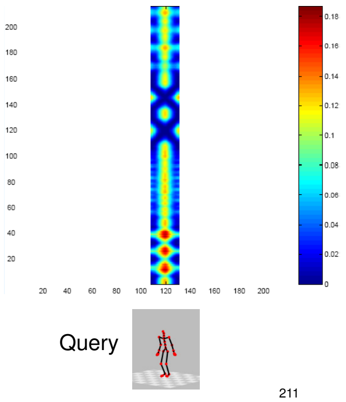




- Given: motion database (one single document)
- Compute: self-similarity matrix

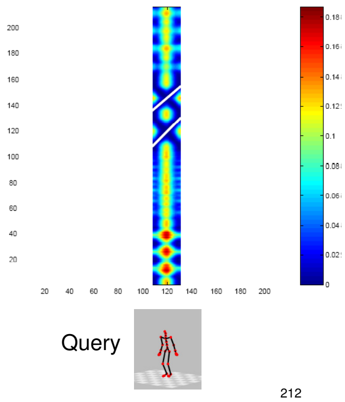
Self-Similarity Matrix

- Query: segment of motion database
- Consider similarity matrix over query



Self-Similarity Matrix

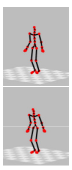
- Identify diagonal paths of low cost

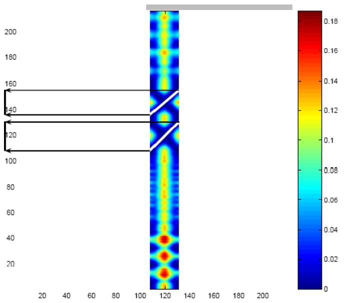


Self-Similarity Matrix

Hit 2

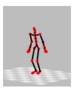
Hit 1





- Identify diagonal paths of low cost
- Project paths onto vertical axis

Query



Some Drawbacks

- DTW-based techniques computationally expensive
  ⇝ do not scale to large databases
- Rely on numerical features
  ⇝ hard to identify logically related motions
- No user-specified "center of attention„
  ⇝ incorporation of a-priori knowledge not possible

Other Recent Approaches

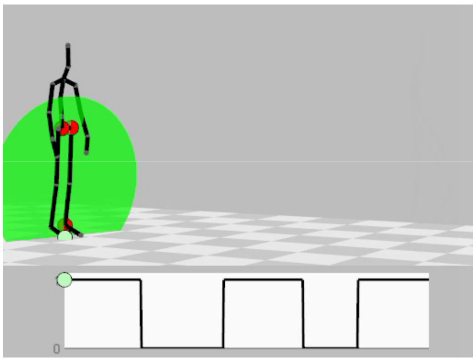
- Wu et al. (IPPR 2003):
  - identify candidates for start and end frames
  - use DTW to compute actual distance from query
- Keogh et al. (VLDB 2004):
  - identify motion clips differing by global scaling
- Forbes/Fiume (SCA 2005):
  - PCA-based local features
  - substring DTW for matching

Our Approach

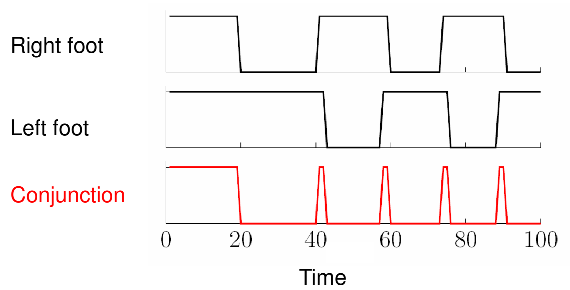
- Introduction of relational features
  ⇝ accounting for spatial deformations
- Introduction of adaptive temporal segmentation
  ⇝ accounting for temporal deformations
- Usage of linear time/space indexing techniques
  ⇝ scalable to large databases

Müller/Röder/Clausen (SIGGRAPH 2005)

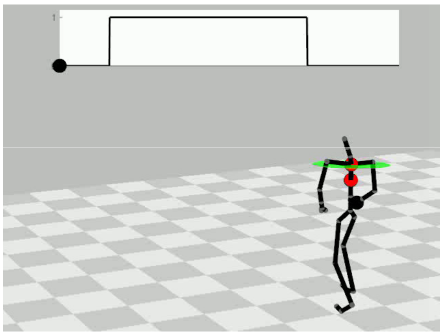
Relational Features



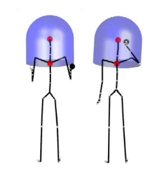
Relational Features



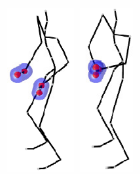
Relational Features



Relational Features

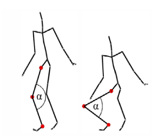


Left hand touching head?

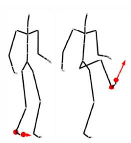


Both hands touching?

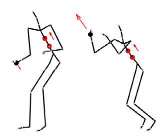
Relational Features



Right knee bent?



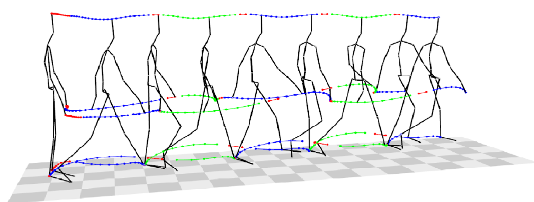
Right foot fast?



Right hand moving upwards?

Relational Features

Temporal Segmentation:

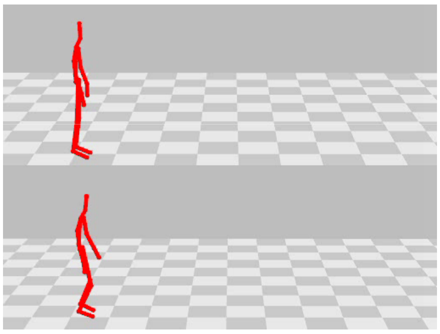


Induced feature sequence:

$$\left(\binom{1}{1}, \binom{0}{1}, \binom{1}{1}, \binom{1}{0}, \binom{1}{1}, \binom{0}{1}, \binom{1}{1}, \binom{1}{0}, \binom{1}{1}, \binom{0}{1}\right)$$

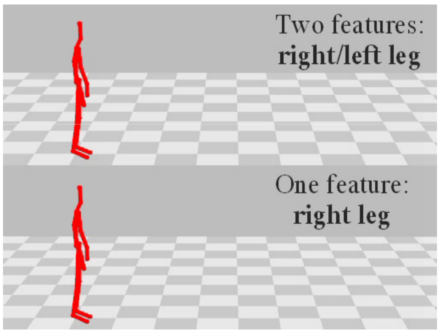
Relational Features

Spatio-temporal invariance

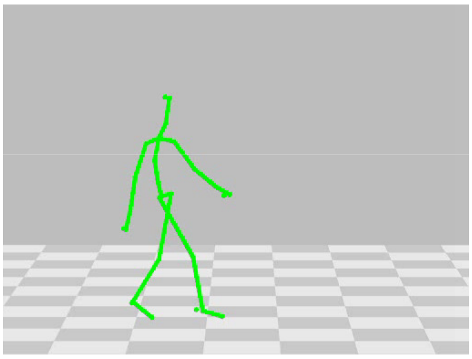


Relational Features

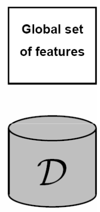
Feature Adaptivity



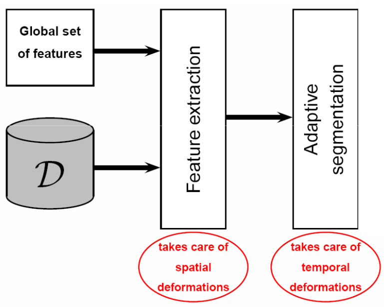
Motion Retrieval



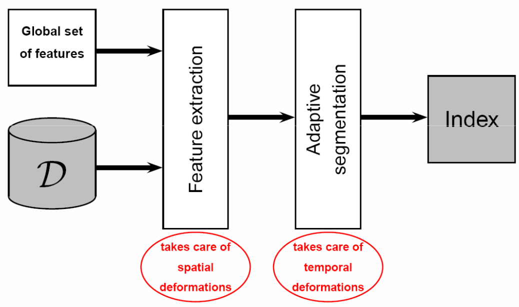
Motion Retrieval



Motion Retrieval

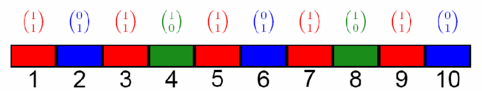


Motion Retrieval



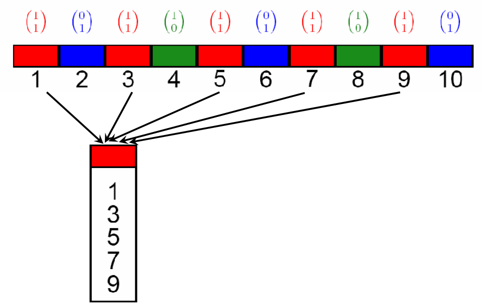
Motion Retrieval

Indexing with inverted lists



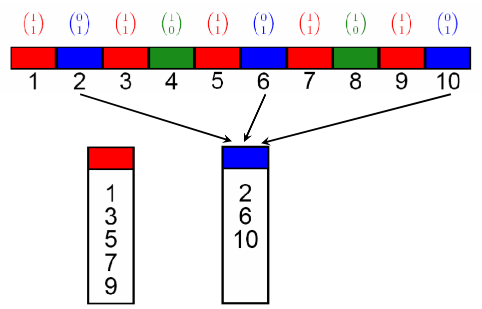
Motion Retrieval

Indexing with inverted lists



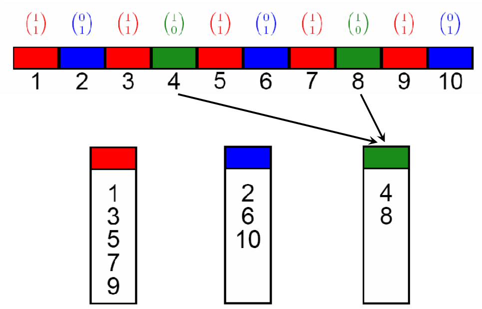
Motion Retrieval

Indexing with inverted lists



Motion Retrieval

Indexing with inverted lists



Motion Retrieval

Preprocessing (Index)

- 3 hours of Mocap data
- 31 (manually designed) boolean featues

Database	Index
1,200,000 frames	230,000 segments
370 MB	7.54 MB

- Index construction: 376 seconds
- Index time and index size linear in #(segments)
- Index is query independent

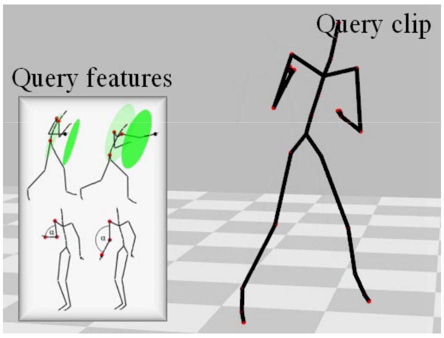
Motion Retrieval

Query and retrieval stage

- Query motion clip
- Optional selection of preferences
  - feature selection
  - degree of fault tolerance
  - ranking strategy
- Automatic conversion of query into feature sequence
- Retrieving hits based on inverted lists
- Typical query response times: 10-300 ms

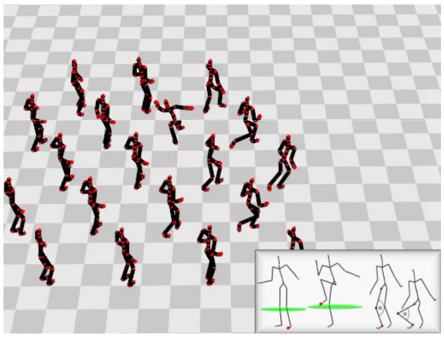
Motion Retrieval

Results: Punch



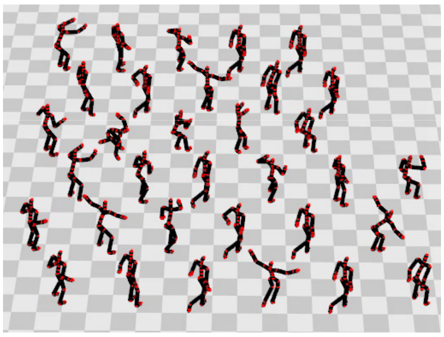
Motion Retrieval

Results: Kick



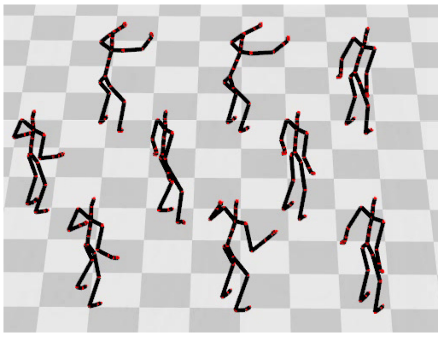
Motion Retrieval

Results: Squat (unranked)



Motion Retrieval

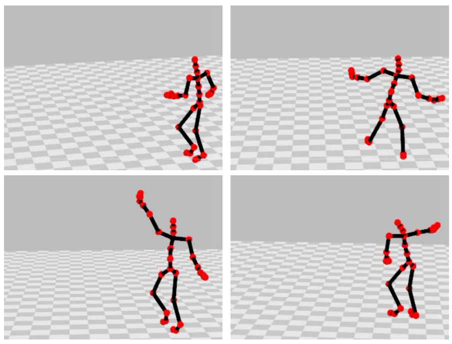
Results: Squat (top 9 ranked)



Strengths and Weaknesses

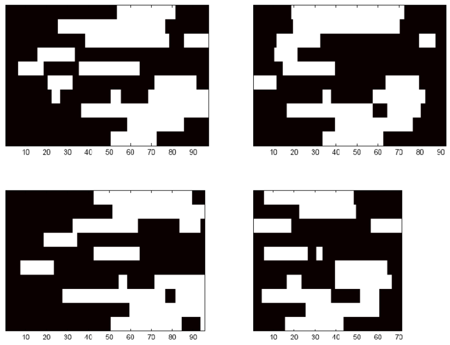
	Strength	Weakness
Retrieval	Efficiency	Rigid False positives/negatives Ranking?
Feature Design	Clear semantics	Ad-hoc Automation?
Feature Selection	A-priori knowledge	Critical Automation

Motion Templates

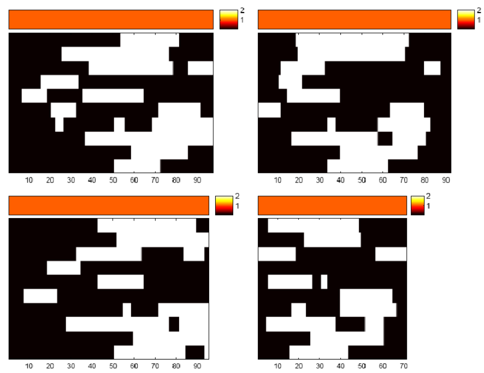


Müller/Röder (SCA 2006)

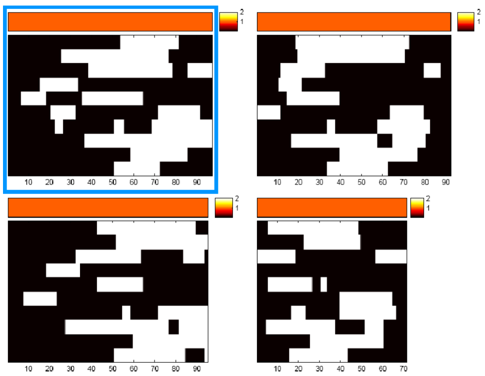
Motion Templates



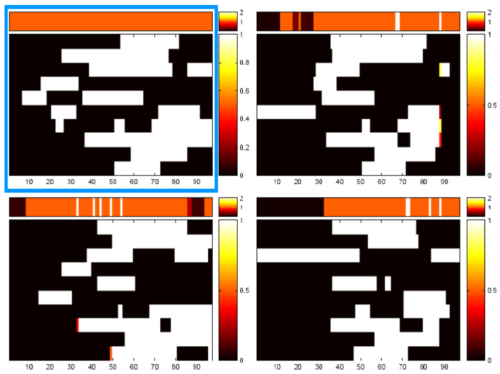
Motion Templates



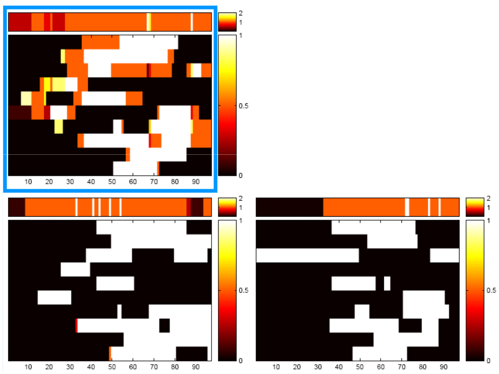
Motion Templates



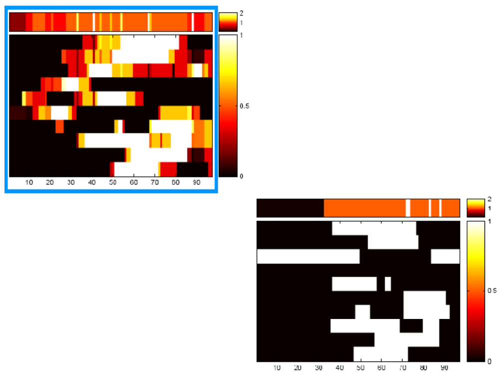
Motion Templates



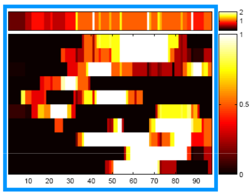
Motion Templates



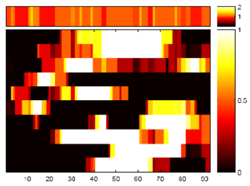
Motion Templates



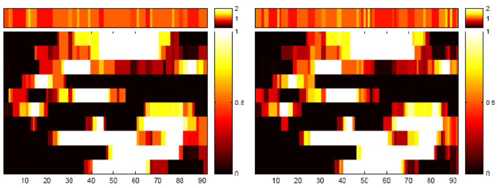
Motion Templates



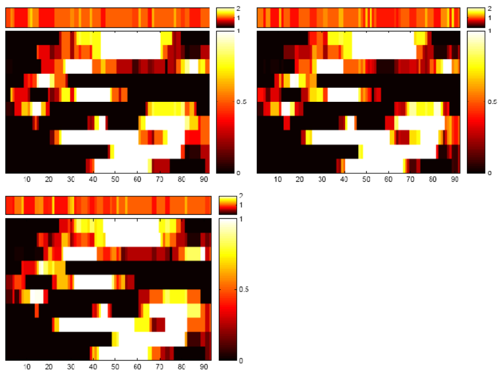
Motion Templates



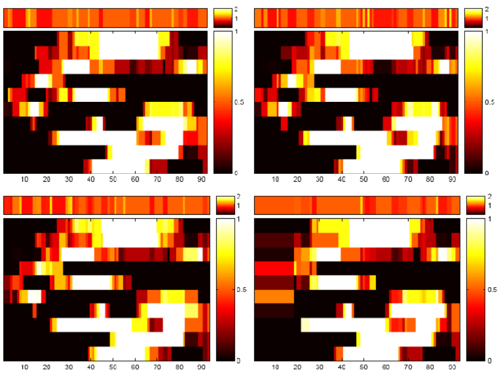
Motion Templates



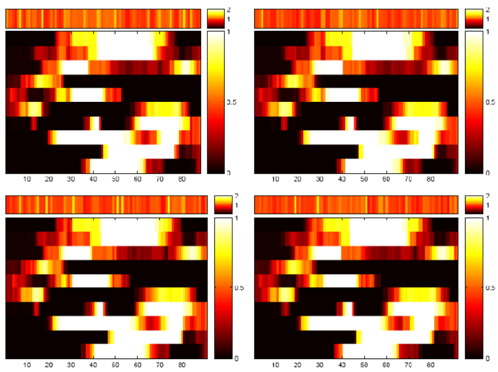
Motion Templates



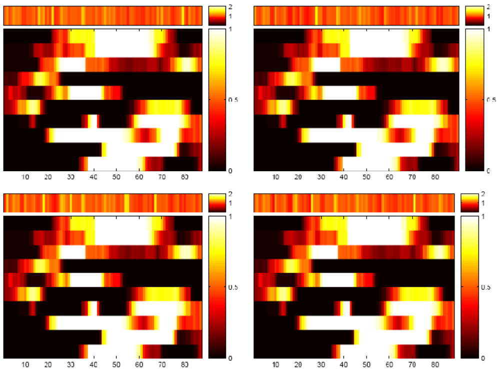
Motion Templates



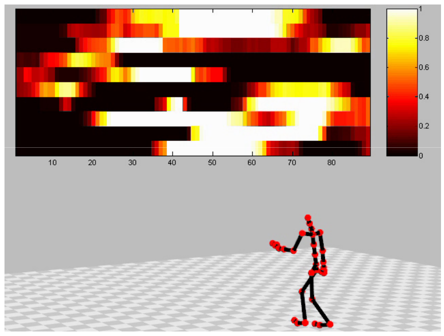
Motion Templates



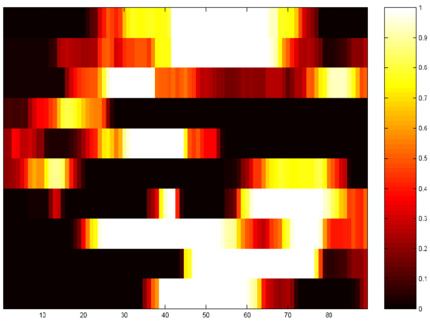
Motion Templates



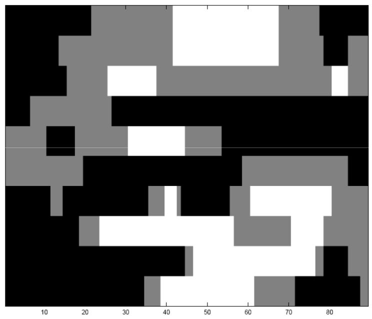
Motion Templates



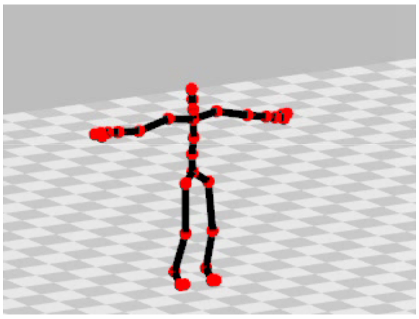
Motion Templates



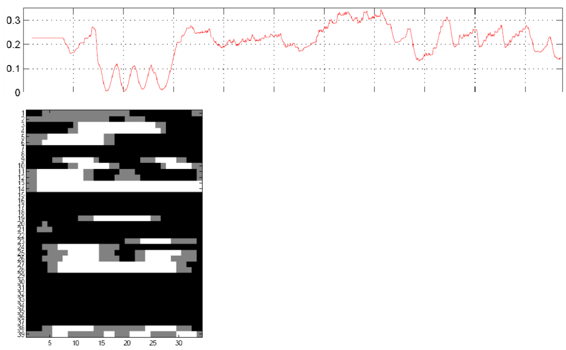
Motion Templates



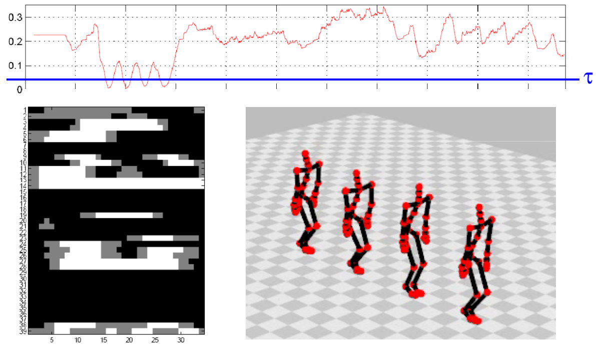
MT-based Motion Retrieval



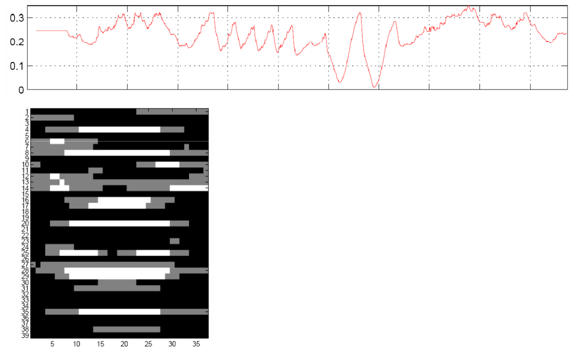
MT-based Motion Retrieval: Jumping Jack



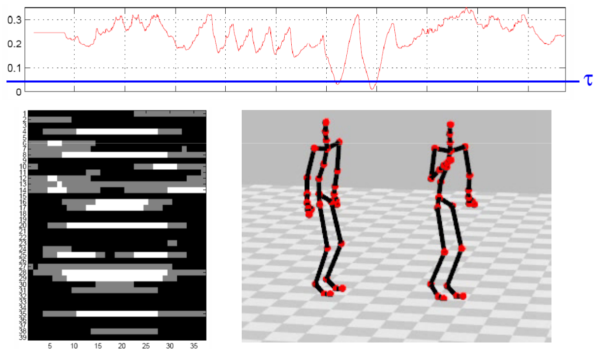
MT-based Motion Retrieval: Jumping Jack



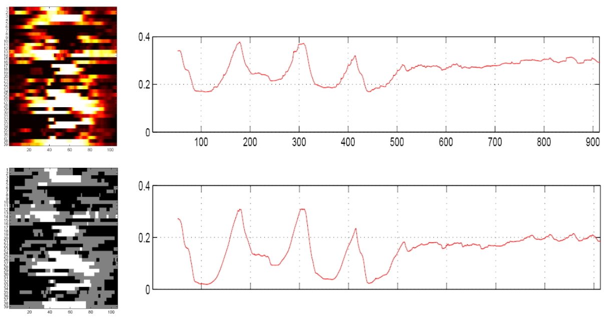
MT-based Motion Retrieval: Elbow-To-Knee



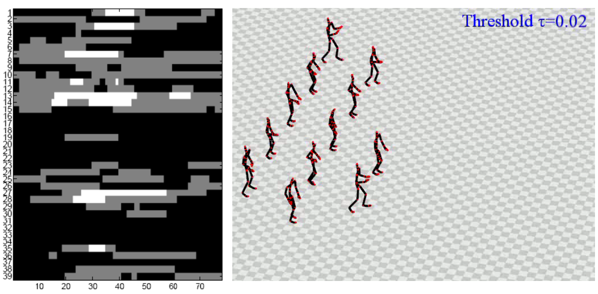
MT-based Motion Retrieval: Elbow-To-Knee



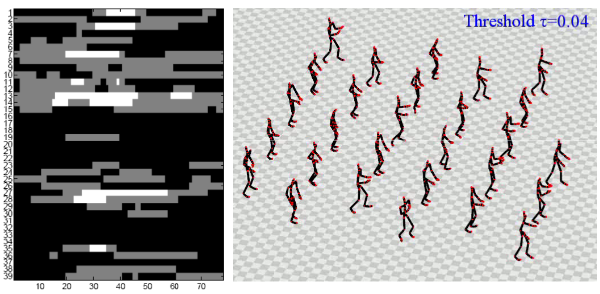
MT-based Motion Retrieval: Cartwheel



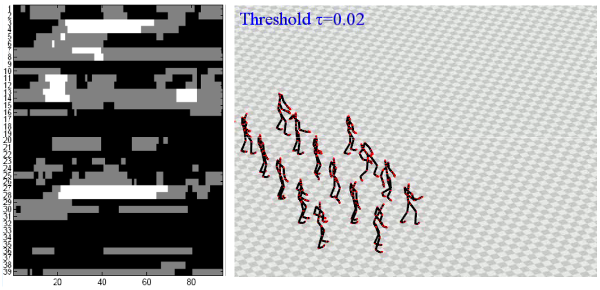
MT-based Motion Retrieval: Throw



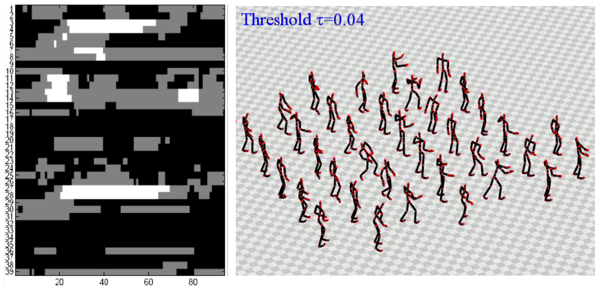
MT-based Motion Retrieval: Throw



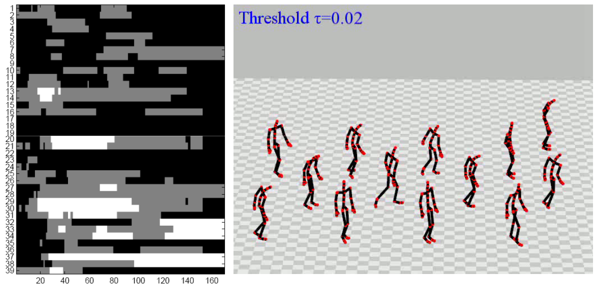
MT-based Motion Retrieval: Basketball



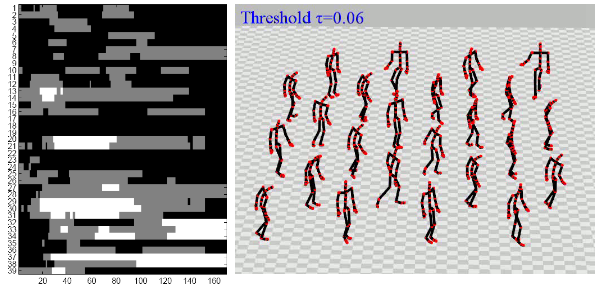
MT-based Motion Retrieval: Basketball



MT-based Motion Retrieval: Lie Down Floor



MT-based Motion Retrieval: Lie Down Floor



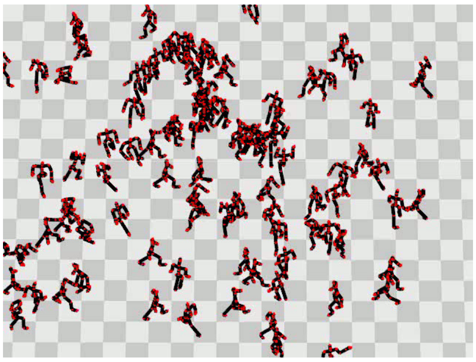
Problems and Future Work

- Efficiency: MT-based matching is linear in database size
- Hit quality: MT-based matching has problems with short motions with few characteristic aspects
- Current work: Combine MT-based matching with aspects of exact matching:
  - “Hard constraints” such as keyframes
  - Index-based preselection

Conclusions

- Automated data organization
- Handling object deformations
- Handling multimodality
- Synchronization (alignment)
- Efficiency

Conclusions



Literature

- Part I: Music Synchronization
- Part II: Audio Structure Analysis
- Part III: Audio Matching
- Part IV: Motion Retrieval

Part I: Music Synchronization

- N. Adams, D. Marquez, and G. H. Wakefield, Iterative deepening for melody alignment and retrieval, in Proc. ISMIR, London, GB, 2005.
- V. Arifi, M. Clausen, F. Kurth, and M. Müller, Synchronization of music data in score-, MIDI- and PCM-format, Computing in Musicology, 13 (2004).
- R. Dannenberg, An on-line algorithm for real-time accompaniment, in Proc. International Computer Music Conference (ICMC), 1984, pp. 193–198.
- R. Dannenberg and N. Hu, Polyphonic audio matching for score following and intelligent audio editors, in Proc. ICMC, San Francisco, USA, 2003, pp. 27–34.
- R. Dannenberg and C. Raphael, Music score alignment and computer accompaniment, Special Issue, Commun. ACM, 49 (2006), pp. 39–43.
- S. Dixon and G. Widmer, Match: A music alignment tool chest, in Proc. ISMIR, London, GB, 2005.
- R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge Univ. Press, 1999.
- C. Fremerey, F. Kurth, M. Müller, and M. Clausen, A demonstration of the SyncPlayer system, in Proc. ISMIR, Vienna, Austria, 2007.

Part I: Music Synchronization

- H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata, and H. Okuno, Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals, ISM, 2006, pp. 257–264.
- L. Grubb and R. Dannenberg, Automated accompaniment of musical ensembles, AAAI, 1994, pp. 94–99.
- N. Hu, R. Dannenberg, and G. Tzanetakis, Polyphonic audio matching and alignment for music retrieval, in Proc. IEEE WASPAA, New Paltz, NY, October 2003.
- F. Kurth, M. Müller, C. Fremerey, Y. Chang, M. Clausen, Automated synchronization of scanned sheet music with audio recordings, in Proc. ISMIR, Vienna, Austria, 2007, pp. 261–266.
- F. Kurth, M. Müller, A. Ribbrock, T. Röder, D. Damm, and C. Fremerey, A prototypical service for real-time access to local context-based music information, in Proc. ISMIR, Barcelona, Spain, 2004.
- M. Müller, D. Appelt, Path-constrained partial music synchronization, in Proc. ICASSP, Las Vegas, USA, 2008.

Part I: Music Synchronization

- M. Müller, F. Kurth, D. Damm, C. Fremerey, and M. Clausen, Lyrics-based audio retrieval and multimodal navigation in music collections, in Proc. ECDL, 2007, pp. 112–123.
- M. Müller, F. Kurth, and T. Röder, Towards an efficient algorithm for automatic score-to-audio synchronization, in Proc. ISMIR, Barcelona, Spain, 2004.
- M. Müller, H. Mattes, and F. Kurth, An efficient multiscale approach to audio synchronization, in Proc. ISMIR, Victoria, Canada, 2006, pp. 192–197.
- N. Orio, Alignment of performances with scores aimed at content-based music access and retrieval, in Proc. ECDL, 2002, pp. 479–492.
- N. Orio, S. Lemouton, D. Schwarz, and N. Schnell, Score following: State of the art and new developments, NIME, 2003, pp. 36–41.
- C. Raphael, A probabilistic expert system for automatic musical accompaniment, Journal of Computational and Graphical Statistics, 10 (2001), pp. 487–512.
- C. Raphael, A hybrid graphical model for aligning polyphonic audio with musical scores, in Proc. ISMIR, Barcelona, Spain, 2004.

Part I: Music Synchronization

- F. Soulez, X. Rodet, and D. Schwarz, Improving polyphonic and polyinstrumental music to score alignment, in Proc. ISMIR, Baltimore, USA, 2003.
- R. J. Turetsky and D. P. Ellis, Force-Aligning MIDI Syntheses for Polyphonic Music Transcription Generation, in Proc. ISMIR, Baltimore, USA, 2003.
- B. Vercoe, The synthetic performer in the context of live performance, in Proc. International Computer Music Conference (ICMC), 1984, pp. 199–200.
- Y. Wang, M.-Y. Kan, T. L. Nwe, A. Shenoy, and J. Yin, Lyrically: Automatic synchronization of acoustic musical signals and textual lyrics, in Proc. ACM Multimedia, New York, USA, 2004, pp. 212–219.

Part II: Audio Structure Analysis

- J. Aucouturier and M. Sandler, Finding repeating patterns in acoustic musical signals, AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, 2002.
- M. A. Bartsch and G. H. Wakefield, To catch a chorus: Using chromabased representations for audio thumbnailing, in Proc. IEEE WASPAA, New Paltz, NY, USA, 2001, pp. 15–18.
- M. A. Bartsch and G. H. Wakefield, Audio thumbnailing of popular music using chroma-based representations, IEEE Trans. on Multimedia, 7 (2005), pp. 96–104.
- W. Chai, Structural analysis of music signals via pattern matching, in Proc. IEEE ICASSP, Hong Kong, China, 2003.
- W. Chai and B. Vercoe, Music thumbnailing via structural analysis, in Proc. ACM Multimedia, 2003.
- M. Cooper and J. Foote, Automatic music summarization via similarity analysis, in Proc. ISMIR, Paris, France, 2002.
- R. Dannenberg and N. Hu, Pattern discovery techniques for music audio, in Proc. ISMIR, Paris, France, 2002.
- J. Foote, Visualizing music and audio using self-similarity, in ACM Multimedia, 1999, pp. 77–80.

Part II: Audio Structure Analysis

- J. Foote, Automatic audio segmentation using a measure of audio novelty, IEEE ICME 2000, pp. 452–455.
- M. Goto, A chorus-section detecting method for musical audio signals, in Proc. IEEE ICASSP, Hong Kong, China, 2003, pp. 437–440.
- M. Goto, SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, in Proc. ACM UIST, 2003, pp. 31–40.
- M. Goto, A chorus section detection method for musical audio signals and its application to a music listening station, IEEE Transactions on Audio, Speech & Language Processing 14 (2006), no. 5, 1783–1794.
- B. Logan and S. Chu, Music summarization using key phrases, in Proc. ICASSP, Istanbul, Turkey, 2000.
- L. Lu, M. Wang, and H.-J. Zhang, Repeating pattern discovery and structure analysis from acoustic music data, in Workshop on Multimedia Information Retrieval, ACM Multimedia, 2004.
- N. C. Maddage, C. Xu, M. S. Kankanhalli, and X. Shao, Content-based music structure analysis with applications to music semantics understanding, in Proc. ACM Multimedia, New York, NY, USA, 2004, pp. 112–119.

Part II: Audio Structure Analysis

- M. Müller and S. Ewert, Joint structure analysis with applications to music annotation and synchronization, to appear in Proc. ISMIR, Philadelphia, USA, 2008.
- M. Müller and F. Kurth, Enhancing similarity matrices for music audio analysis, in Proc. IEEE ICASSP, Toulouse, France, 2006.
- M. Müller and F. Kurth, Towards structural analysis of audio recordings in the presence of musical variations, EURASIP Journal on Advances in Signal Processing, Article ID 89686 (2007).
- G. Peeters, Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach, Proc. ISMIR, Vienna, Austria, 2007.
- G. Peeters, A. L. Burthe, and X. Rodet, Toward automatic music audio summary generation from signal analysis, in Proc. ISMIR, Paris, France, 2002.
- C. Rhodes, M. Casey, Algorithms for determining and labelling approximate hierarchical self-similarity, Proc. ISMIR, Vienna, Austria, 2007.
- C. Xu, N. Maddage, and X. Shao, Automatic music classification and summarization, IEEE Trans. on Speech and Audio Processing, 13 (2005), pp. 441–450.

Part III: Audio Matching

- E. Allamanche, J. Herre, B. Fröba, and M. Cremer, AudioID: Towards Content-Based Identification of Audio Material, in Proc. 110th AES Convention, Amsterdam, NL, 2001.
- P. Cano, E. Battle, T. Kalker, and J. Haitsma, A Review of Audio Fingerprinting, in Proc. 5. IEEE MMSP, St. Thomas, Virgin Islands, USA, 2002.
- M. Casey and M. Slaney, Song intersection by approximate nearest neighbor search, Proc. ISMIR, Victoria, Canada, 2006, pp. 144–149.
- E. Gómez and P. Herrera, The song remains the same: identifying versions of the same piece using tonal descriptors, in Proc. ISMIR, Victoria, Canada, 2006, pp. 180–185.
- J. Haitsma and T. Kalker, A highly robust audio fingerprinting system, in Proc. ISMIR, Paris, France, 2002.
- C. Fremerey, M. Müller, F. Kurth, M. Clausen, Automatic mapping of scanned sheet music to audio recordings, to appear in Proc. ISMIR, Philadelphia, USA, 2008.

Part III: Audio Matching

- F. Kurth, M. Clausen, and A. Ribbrock, Identification of highly distorted audio material for querying large scale data bases, in Proc. 112th AES Convention, Munich, Germany, 2002.
- F. Kurth, M. Müller, Efficient Index-based Audio Matching. IEEE Trans. on Audio, Speech, and Language Processing 16(2) (2008) 382–395.
- M. Müller, F. Kurth, and M. Clausen, Audio matching via chroma-based statistical features, in Proc. ISMIR, London, GB, 2005.
- J. Pickens, J. P. Bello, G. Monti, T. Crawford, M. Dovey, M. Sandler, and D. Byrd, Polyphonic score retrieval using polyphonic audio, in Proc. ISMIR, Paris, 2002.
- J. Serrà and E. Gómez, Audio cover song identification based on tonal sequence alignment , in Proc. IEEE ICASSP, 2008, pp. 61-64.
- P. Shrestha and T. Kalker, Audio fingerprinting in peer-to-peer networks, in Proc. ISMIR, Barcelona, Spain, 2004.
- A. Wang, An industrial strength audio search algorithm, in Proc. ISMIR, Baltimore, USA, 2003.

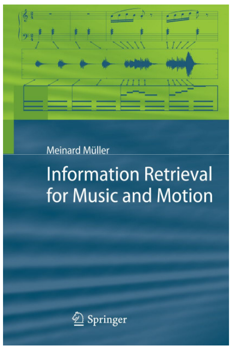
Part IV: Motion Retrieval

- CMU, Carnegie-Mellon Mocap Database. http://mocap.cs.cmu.edu, 2003.
- K. Forbes and E. Fiume, An efficient search algorithm for motion data using weighted PCA, in Proc. 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM Press, 2005, pp. 67–76.
- E. J. Keogh, T. Palpanas, V. B. Zordan, D. Gunopulos, and M. Cardle, Indexing large human-motion databases, in Proc. 30th VLDB Conf., Toronto, 2004, pp. 780–791.
- L. Kovar and M. Gleicher, Automated extraction and parameterization of motions in large data sets, ACM Trans. Graph., 23 (2004), pp. 559–568.
- G. Liu, J. Zhang, W. Wang, and L. McMillan, A system for analyzing and indexing human-motion databases, in Proc. 2005 ACM SIGMOD Intl. Conf. on Management of Data, ACM Press, 2005, pp. 924–926.
- M. Müller and T. Röder, Motion templates for automatic classification and retrieval of motion capture data, in SCA '06: Proc. 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM Press, 2006, pp. 137–146.
- M. Müller and T. Röder, A relational approach to content-based analysis of motion capture data, in Human Motion—Understanding, Modeling, Capture and Animation, B. Rosenhahn, R. Klette, and D. Metaxas, eds., Springer, Berlin, 2007.

Part IV: Motion Retrieval

- M. Müller, T. Röder, and M. Clausen, Efficient content-based retrieval of motion capture data, ACM Trans. Graph., 24 (2005), pp. 677–685.
- A. Witkin and Z. Popović, Motion warping, in Proc. ACM SIGGRAPH 95, Computer Graphics Proc., ACM Press/ACM SIGGRAPH, 1995, pp. 105–108.
- M.-Y.Wu, S. Chao, S. Yang, and H. Lin, Content-based retrieval for human motion data, in 16th IPPR Conf. on Computer Vision, Graphics and Image Processing, 2003, pp. 605–612.
- M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, Documentation of the macoap database HDM05. Computer Graphics Technical Report, CG-2007-2, Department of Computer Science II, University of Bonn, 2007.
- K. Pullen and C. Bregler, Motion capture assisted animation: Texturing and synthesis, ACM Trans. Graph., (2002), pp. 501–508.
- Y. Sakamoto, S. Kuriyama, and T. Kaneko, Motion map: image based retrieval and segmentation of motion data, in Proc. 2004 ACM SIGGRAPH/ Eurographics Symposium on Computer Animation, ACM Press, 2004, pp. 259–266.

Book



Müller, Meinard
Information Retrieval for Music and Motion
2007, XVI. 318 pages
136 illus. 39 in Color, Hardcover
ISBN: 978-3-540-74047-6
www.springer.com/978-3-540-74047-6/
69,50 EUR