INTERNATIONAL AUDIO LABORATORIES ERLANGEN

AUDIO LABS

Informed Feature Representations for Music and Motion

Meinard Müller

Lorentz Workshop

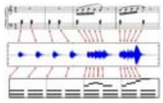
Music Similarity: Concepts, Cognition and Computation

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Meinard Müller

- 2007 Habilitation, Bonn
- 2007 MPI Informatik, Saarbrücken
  Senior Researcher
  Music Processing & Motion Processing

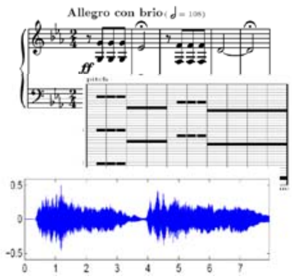


- 2012 W3-Professur, AudioLabs Erlangen
  Semantic Audio Processing

Thanks

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

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

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

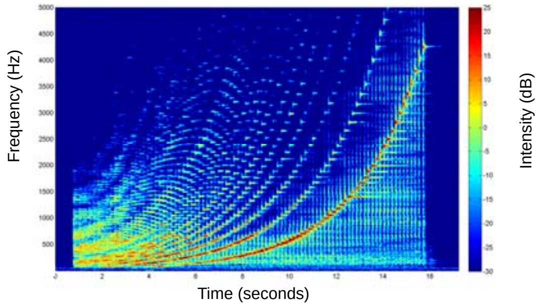
Chroma-based Audio Features

- 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

Chroma-based Audio Features

Example: Chromatic scale

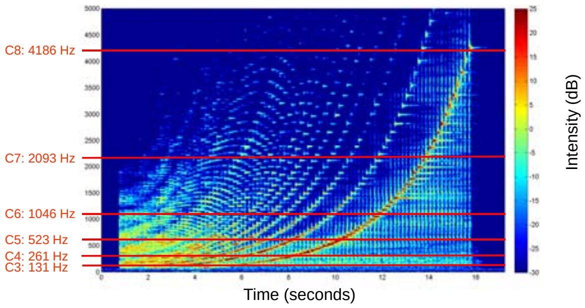
Spectrogram



Chroma-based Audio Features

Example: Chromatic scale

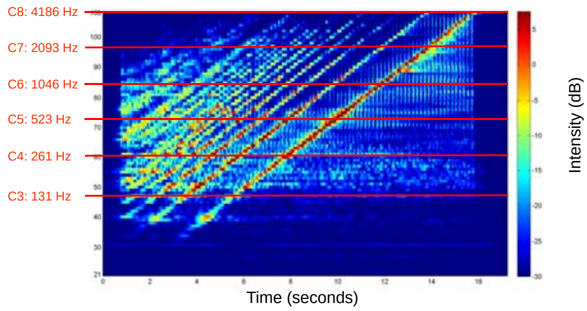
Spectrogram



Chroma-based Audio Features

Example: Chromatic scale

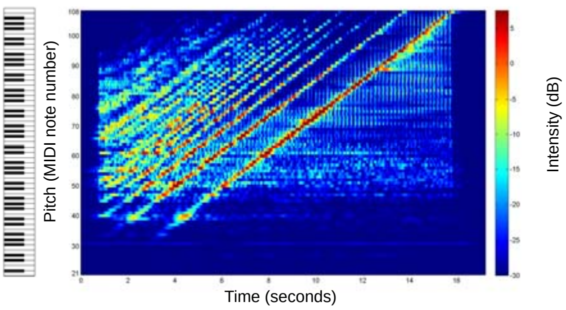
Log-frequency spectrogram



Chroma-based Audio Features

Example: Chromatic scale

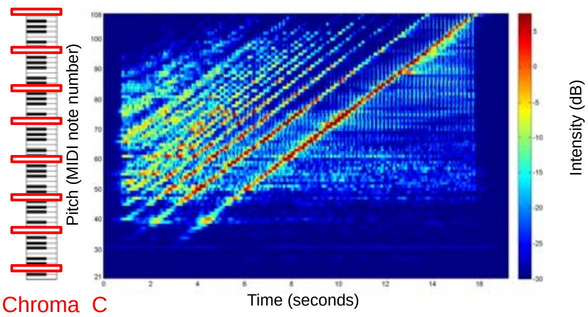
Log-frequency spectrogram



Chroma-based Audio Features

Example: Chromatic scale

Log-frequency spectrogram

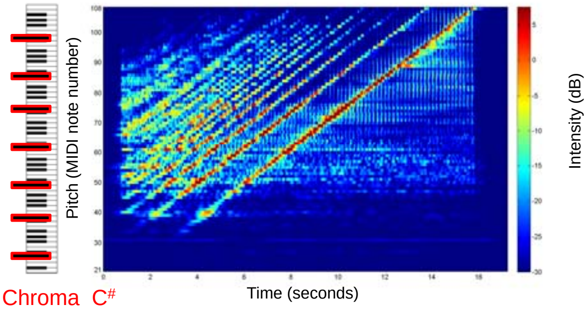


Chroma C

Chroma-based Audio Features

Example: Chromatic scale

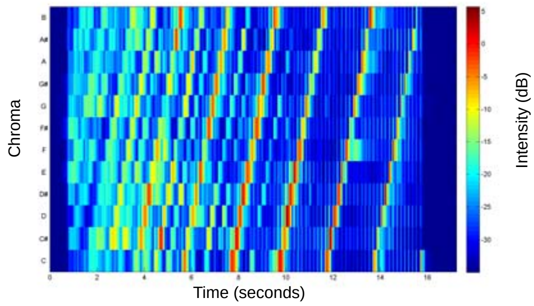
Log-frequency spectrogram



Chroma-based Audio Features

Example: Chromatic scale

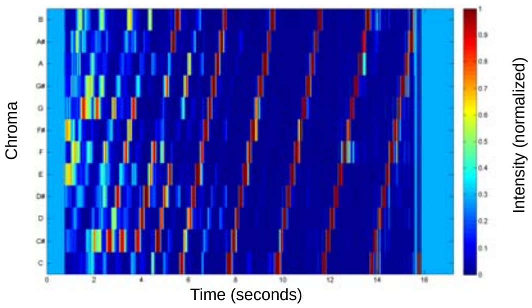
Chroma representation



Chroma-based Audio Features

Example: Chromatic scale

Chroma representation (normalized, Euclidean)

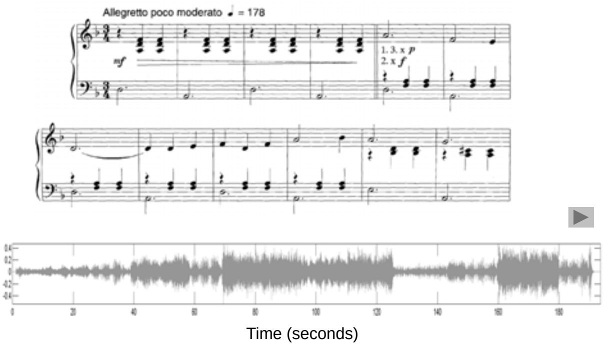


Enhancing Chroma Features

- 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

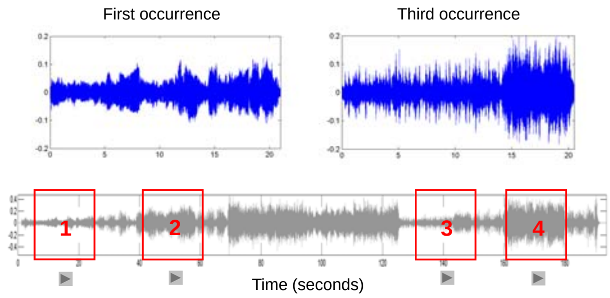
M. Müller and S. Ewert
Towards Timbre-Invariant Audio Features for Harmony-Based Music.
IEEE Trans. on Audio, Speech & Language Processing, Vol. 18, No. 3, pp. 649-662, 2010.

Motivation: Audio Matching

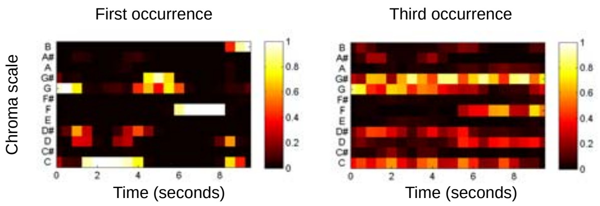


Motivation: Audio Matching

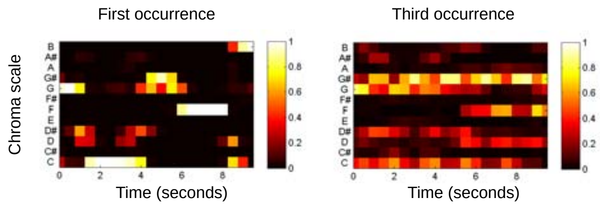
Four occurrences of the main theme



Chroma Features

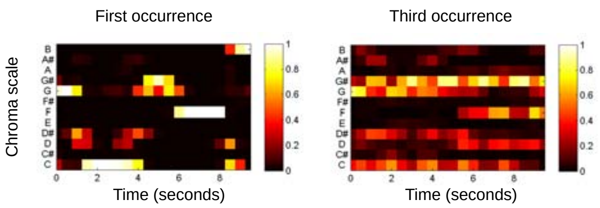


Chroma Features



How to make chroma features more robust to timbre changes**?**

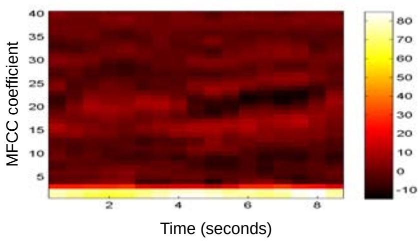
Chroma Features



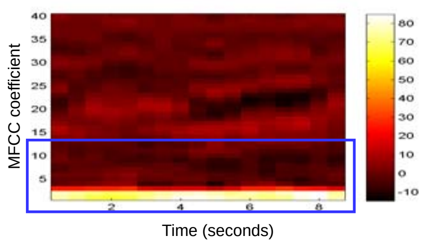
How to make chroma features more robust to timbre changes**?**

Idea: Discard timbre-related information

MFCC Features and Timbre

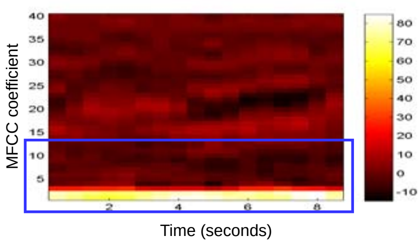


MFCC Features and Timbre



Lower MFCCs ⟷ Timbre

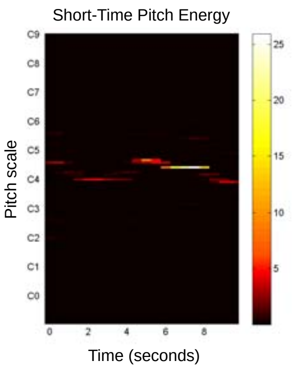
MFCC Features and Timbre



Lower MFCCs ⟷ Timbre

Idea: Discard lower MFCCs to achieve timbre invariance

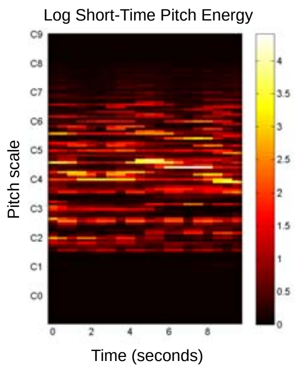
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram

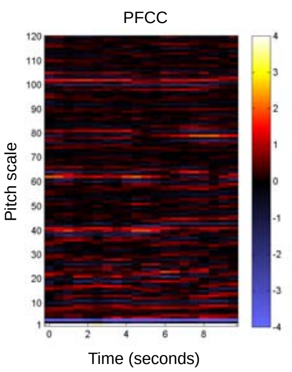
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)

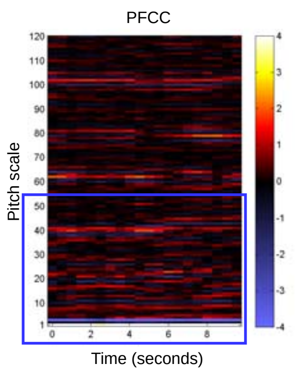
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT

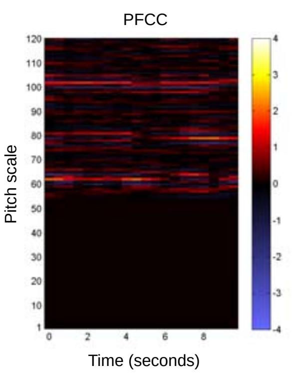
Enhancing Timbre Invariance



Steps:

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

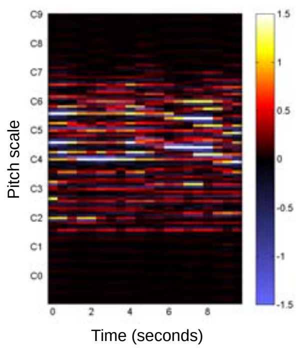
Enhancing Timbre Invariance



Steps:

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

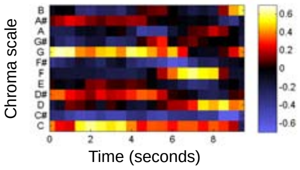
Enhancing Timbre Invariance



Steps:

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

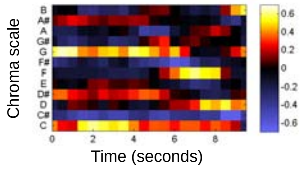
Enhancing Timbre Invariance



Steps:

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

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

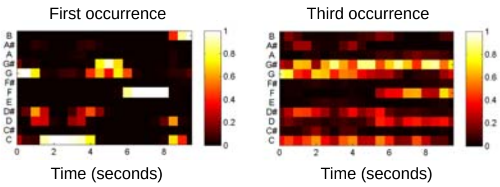
CRP(n)

Chroma DCT-**R**educed Log-**P**itch

Chroma versus CRP

Shostakovich Waltz

Chroma

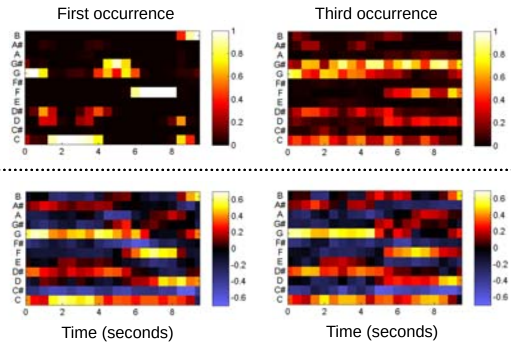


Chroma versus CRP

Shostakovich Waltz

Chroma

CRP(55)
n = 55



Audio Analysis

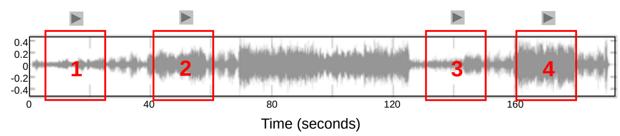
Idea:

Use “Audio Matching” for analyzing and understanding audio & feature properties:

- Relative comparison
- Compact
- Intuitive
- Quantitative evaluation

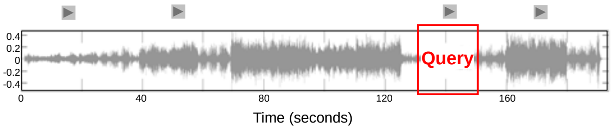
Audio Analysis

Example: Shostakovich, Waltz (Yablonsky)



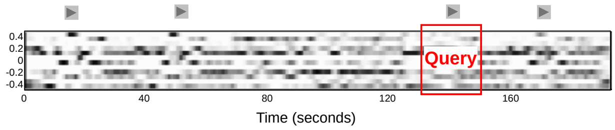
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



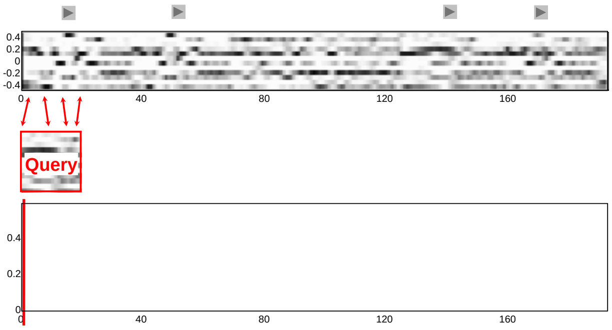
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



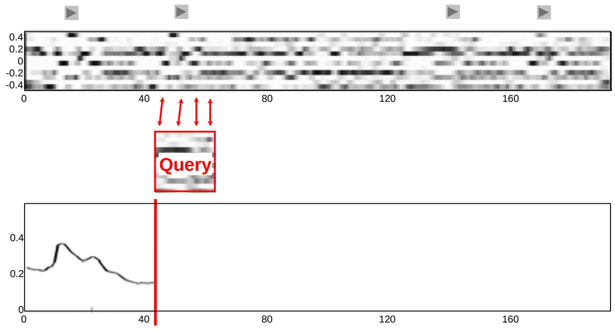
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



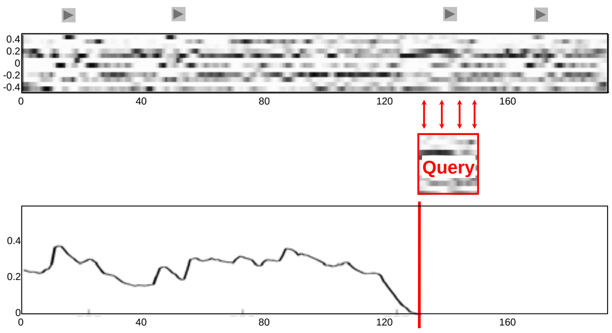
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



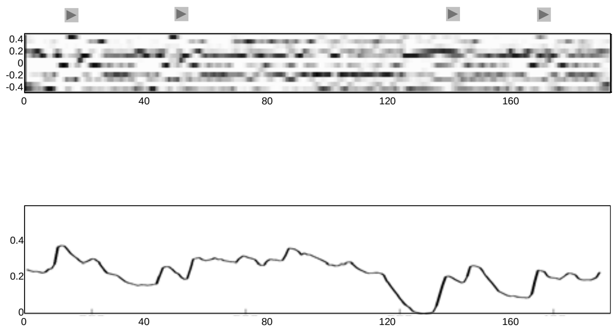
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



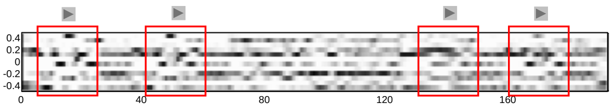
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)

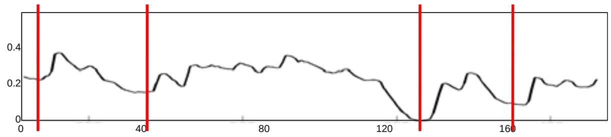


Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



Expected matching positions (should have local minima)

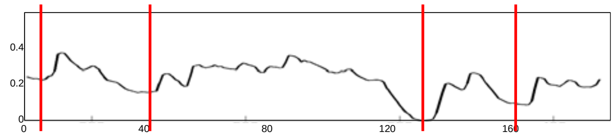


Audio Analysis

Idea:

- Use matching curve for analyzing feature properties

Expected matching positions (should have local minima)

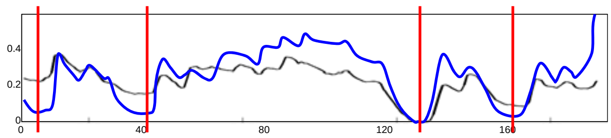


Audio Analysis

Idea:

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

Expected matching positions (should have local minima)



Quality: Audio Matching

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

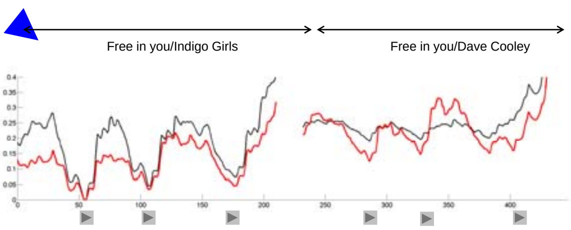
Standard Chroma (Chroma Pitch)

Time (seconds)

Quality: Audio Matching

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

Standard Chroma (Chroma Pitch)
CRP(55)



Time (seconds)

Chroma Toolbox

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

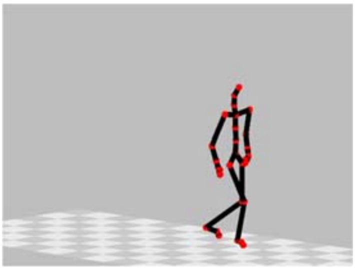
- http://www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/
- MATLAB implementations for various chroma variants

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

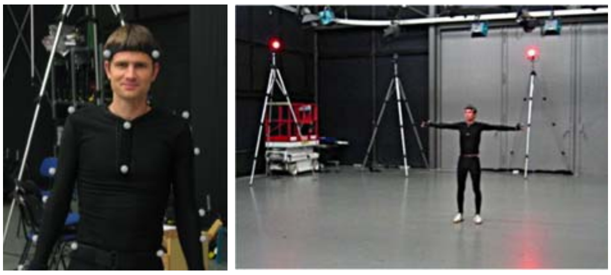
Motion Capture Data

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

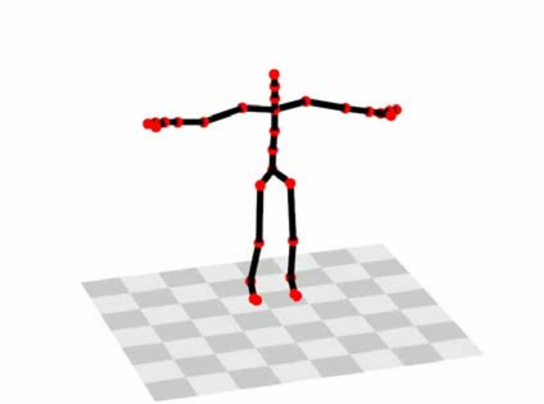


Motion Capture Data

Optical System

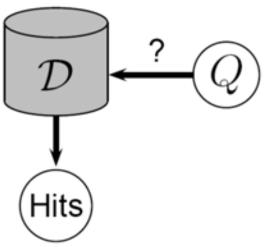


Motion Capture Data

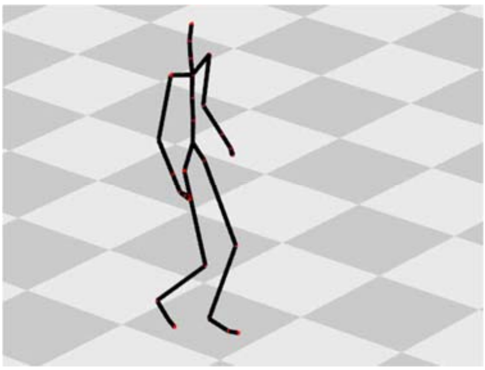


Motion Retrieval

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

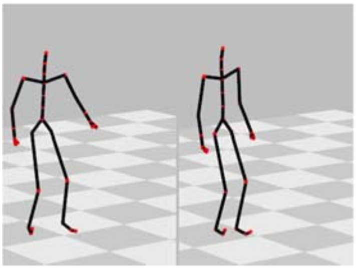


Motion Retrieval



Motion Retrieval

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



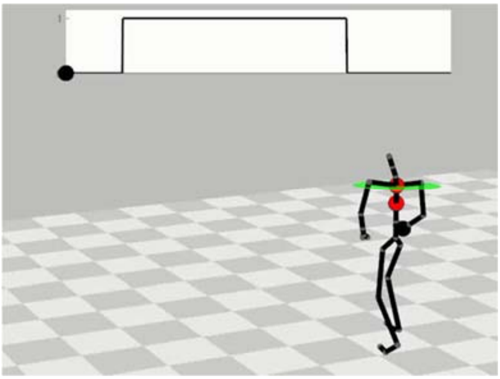
Relational Features

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

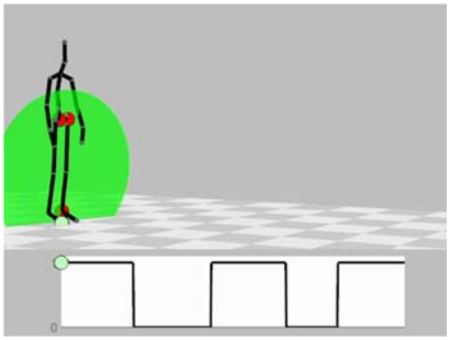
Meinard Müller, Tido Röder, and Michael Clausen
Efficient content-based retrieval of motion capture data.
ACM Transactions on Graphics (SIGGRAPH), vol. 24, pp. 677-685, 2005.

Meinard Müller and Tido Röder
Motion templates for automatic classification and retrieval of motion capture data.
Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA), Vienna, Austria, pp. 137-146, 2006.

Relational Features



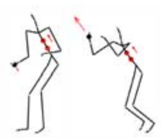
Relational Features



Relational Features

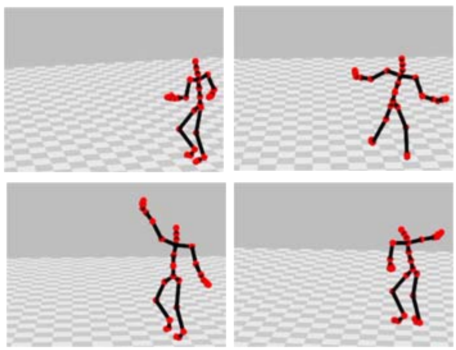
Right knee bent?

Right foot fast?

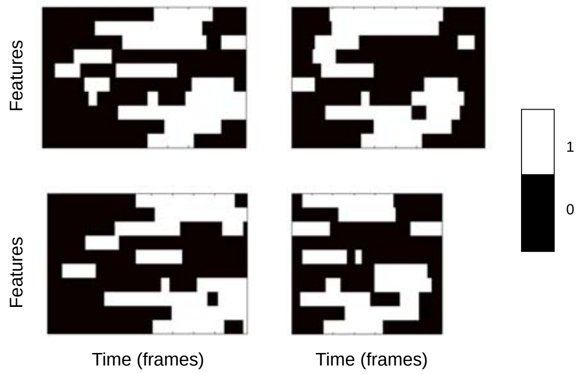


Right hand moving upwards?

Motion Templates (MT)

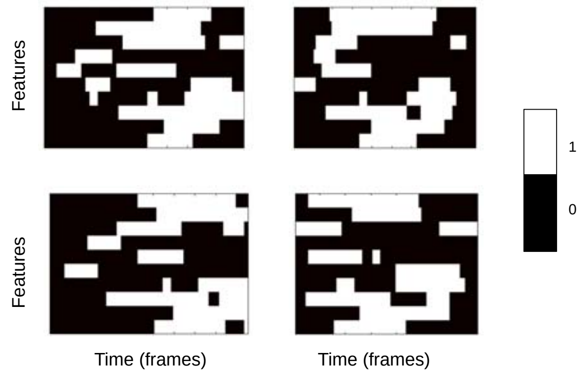


Motion Templates (MT)



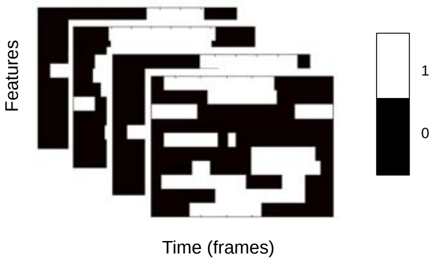
Motion Templates (MT)

Temporal alignment



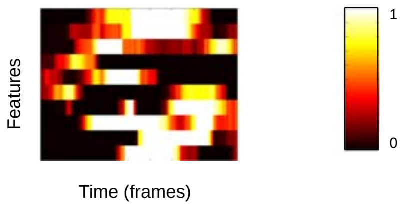
Motion Templates (MT)

Superimpose templates

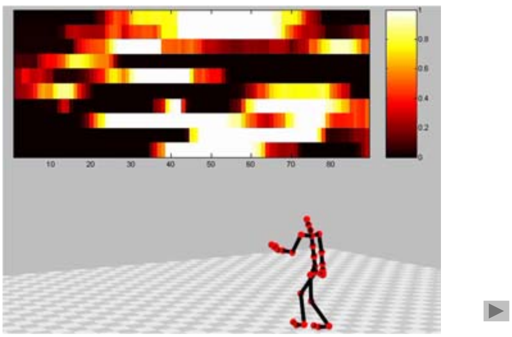


Motion Templates (MT)

Compute average

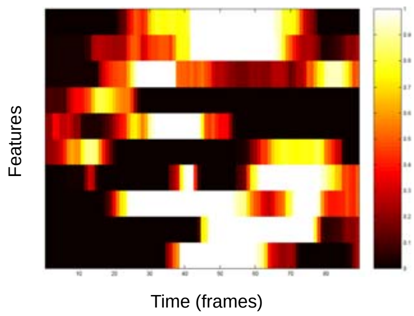


Motion Templates (MT)



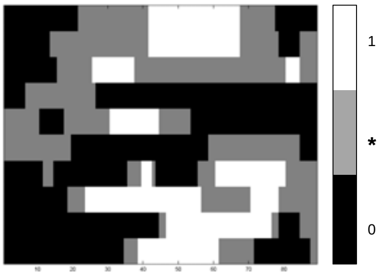
Motion Templates (MT)

Average template



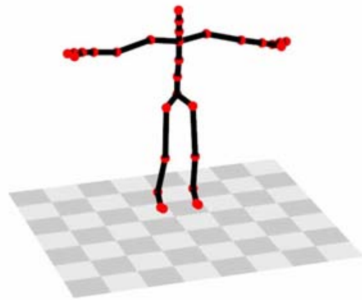
Motion Templates (MT)

Quantized template

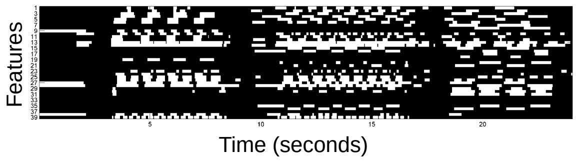


- Gray areas indicate inconsistencies / variations
- Achieve invariance by disregarding gray areas

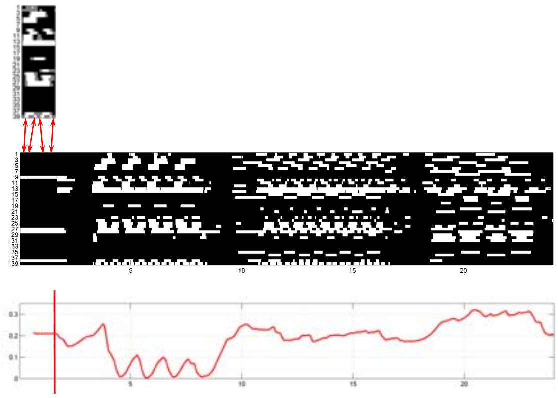
MT-based Motion Retrieval



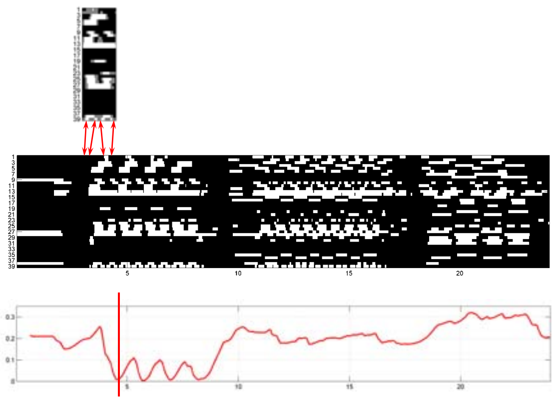
MT-based Motion Retrieval



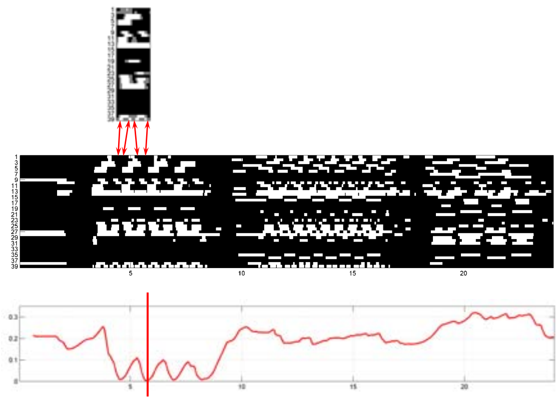
MT-based Motion Retrieval: Jumping Jack



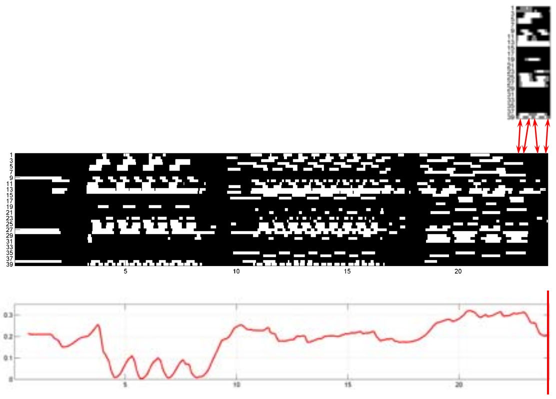
MT-based Motion Retrieval: Jumping Jack



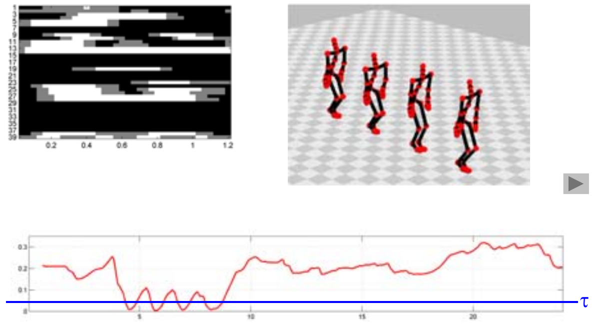
MT-based Motion Retrieval: Jumping Jack



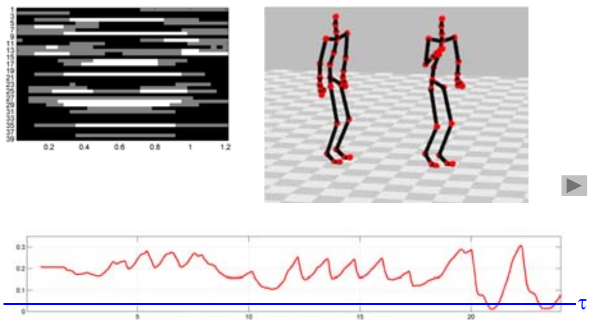
MT-based Motion Retrieval: Jumping Jack



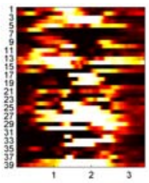
MT-based Motion Retrieval: Jumping Jack



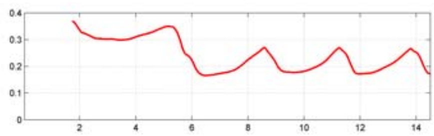
MT-based Motion Retrieval: Elbow-To-Knee

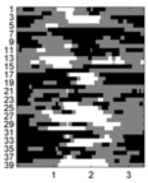


MT-based Motion Retrieval: Cartwheel

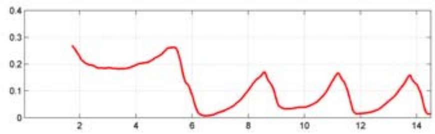


Matching curve using average MT

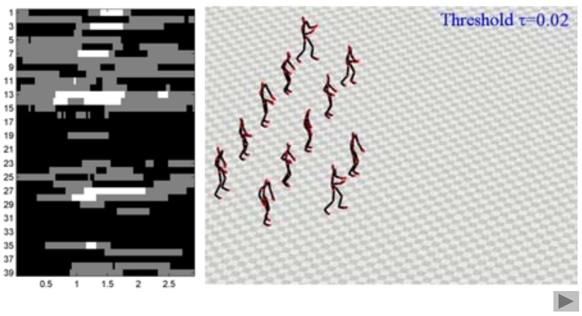




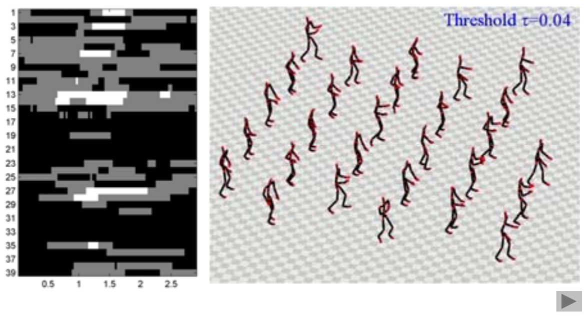
Matching curve blending out variations



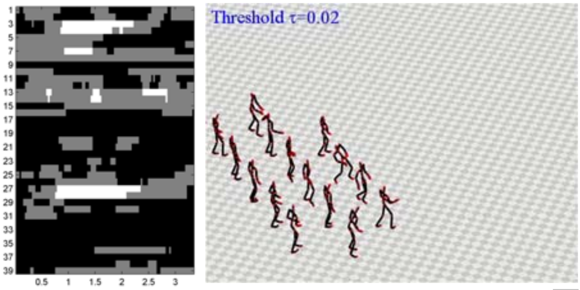
MT-based Motion Retrieval: Throw



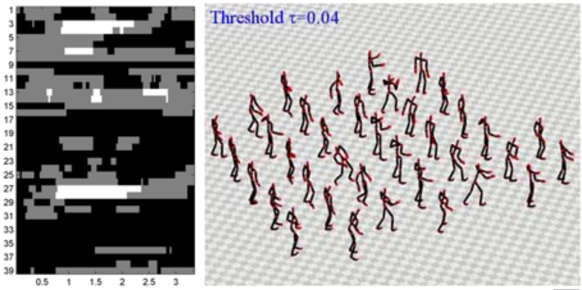
MT-based Motion Retrieval: Throw



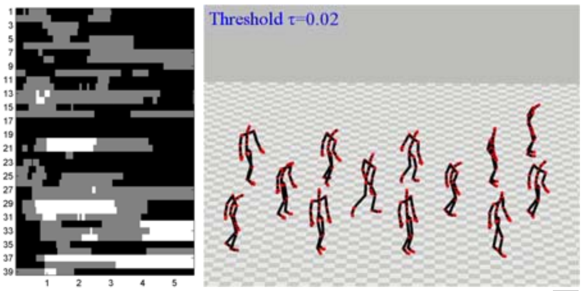
MT-based Motion Retrieval: Basketball



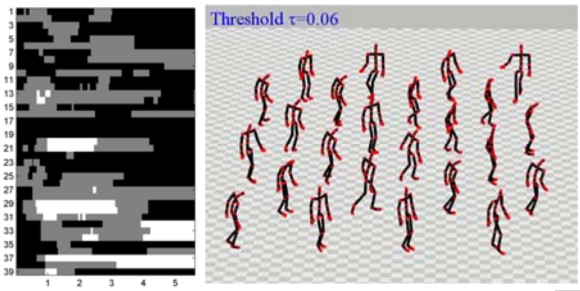
MT-based Motion Retrieval: Basketball



MT-based Motion Retrieval: Lie Down Floor



MT-based Motion Retrieval: Lie Down Floor



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

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
- …

Music Signal Processing

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

Relative comparison of music audio data

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

Music Signal Processing

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

Relative comparison of music audio data

Need of robust mid-level representations

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

Mid-Level Representations

Musical Aspect	Features	Dimension
Timbre	MFCC features	10 - 15
Harmony	Pitch features	60 - 120
Harmony	Chroma features	12
Tempo	Tempogram	> 100

Mid-Level Representations

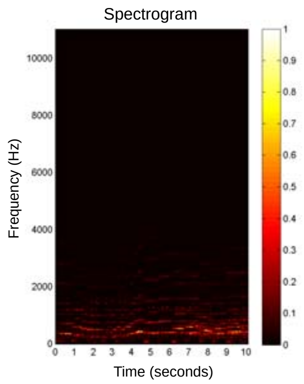
Musical Aspect	Features	Dimension
Timbre	MFCC features	10 - 15
Harmony	Pitch features	60 - 120
Harmony	Chroma features	12
Tempo	Tempogram	> 100
Tempo	Cyclic tempogram	10 - 30

Peter Grosche, Meinard Müller, and Frank Kurth
Cyclic tempogram – a mid-level tempo representation for music signals.
Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, Texas, USA, pp. 5522-5525, 2010.

Novelty Curve

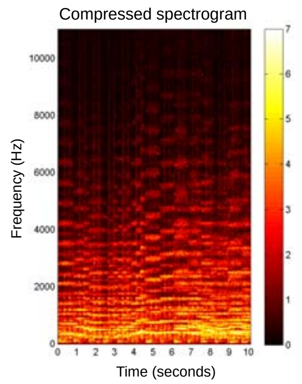
Example: Waltz, Jazz Suite No. 2

Novelty Curve



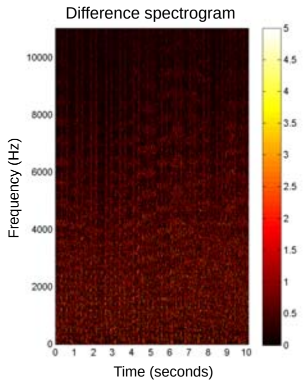
Steps:
1. Spectrogram

Novelty Curve



Steps:
1. Spectrogram
2. Log compression

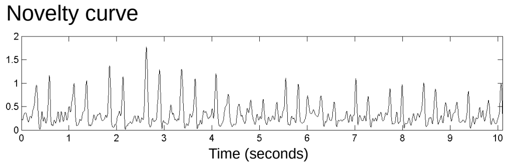
Novelty Curve



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

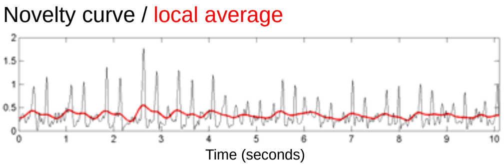
Novelty Curve

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



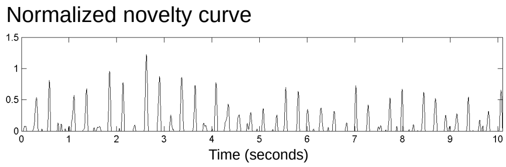
Novelty Curve

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

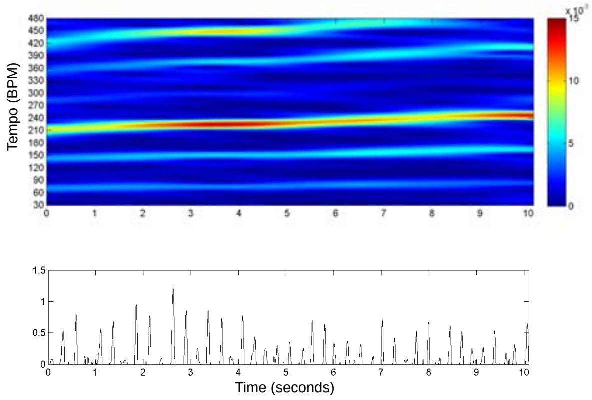


Novelty Curve

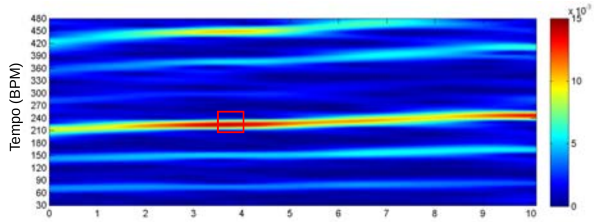
Steps:
1. Spectrogram
2. Log compression
3. Differentiation
4. Accumulation
5. Normalization



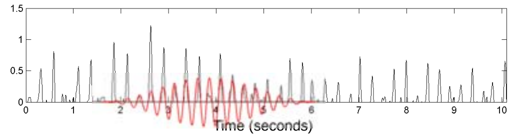
Tempogram



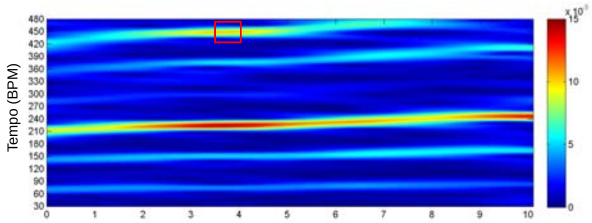
Tempogram



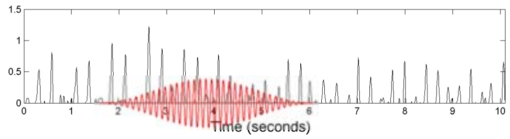
Short-time Fourier analysis



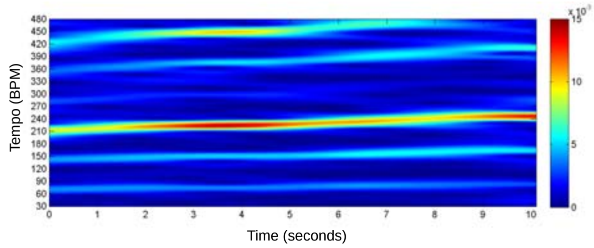
Tempogram



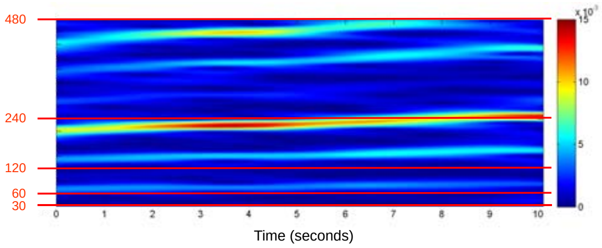
Short-time Fourier analysis



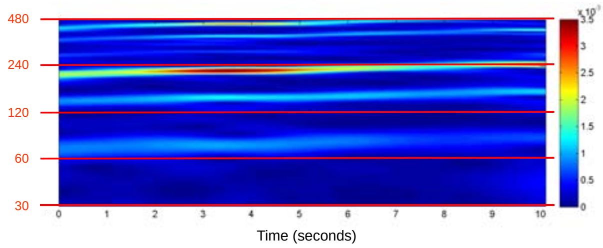
Tempogram



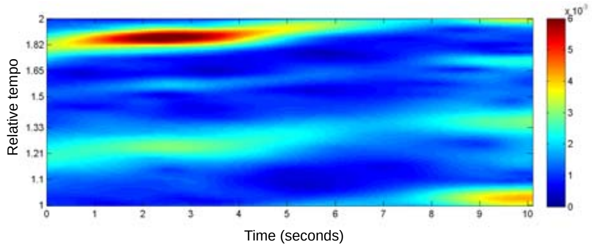
Tempogram



Log-Scale Tempogram



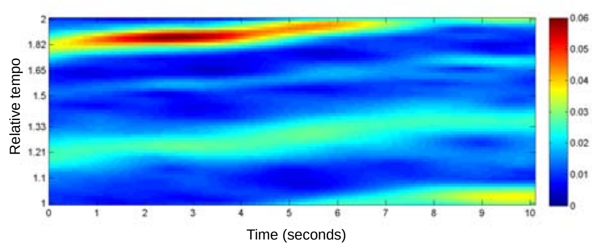
Cyclic Tempogram



Cylic projection

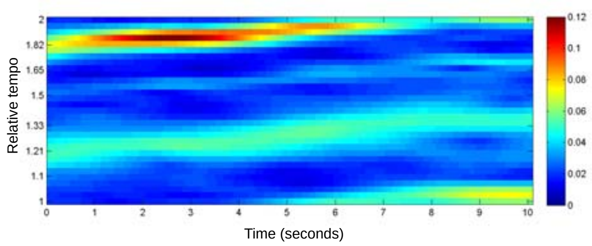
Relative to tempo class […,30,60,120,240,480,…]

Cyclic Tempogram



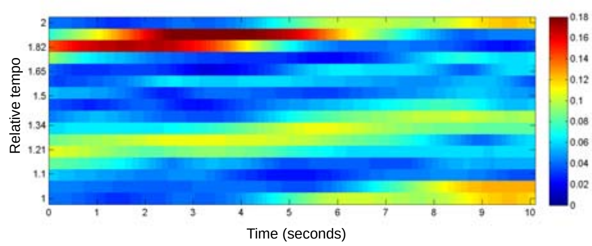
Quantization: 60 tempo bins

Cyclic Tempogram



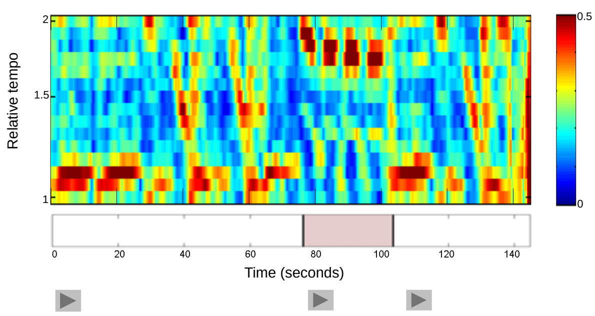
Quantization: 30 tempo bins

Cyclic Tempogram



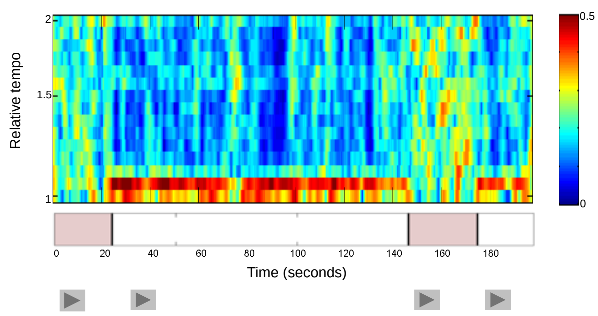
Quantization: 15 tempo bins

Audio Segmentation



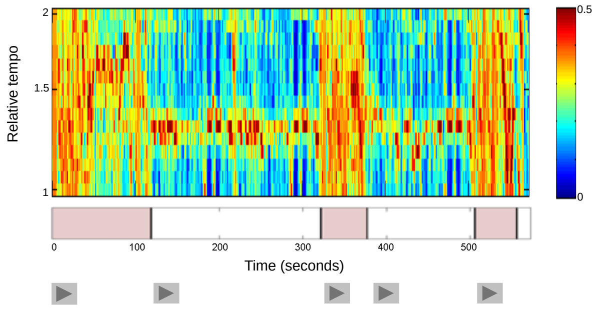
Example: Brahms Hungarian Dance No. 5

Audio Segmentation



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

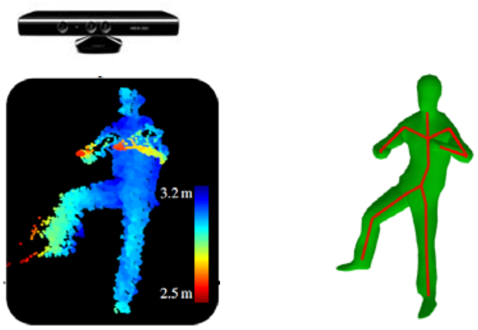
Data-Driven Motion Reconstruction

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

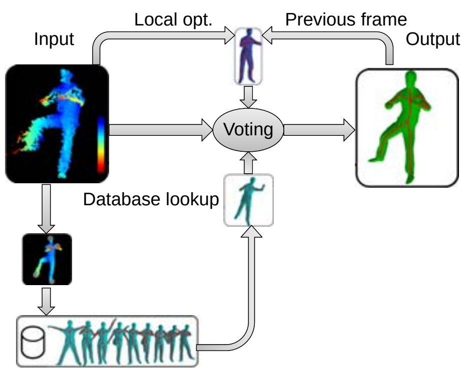
Andreas Baak, Meinard Müller, Gaurav Bharaj, Hans-Peter Seidel, and Christian Theobalt
A data-driven approach for real-time full body pose reconstruction from a depth camera.
Proceedings of the 13th International Conference on Computer Vision (ICCV), 2011.

Data-Driven Motion Reconstruction

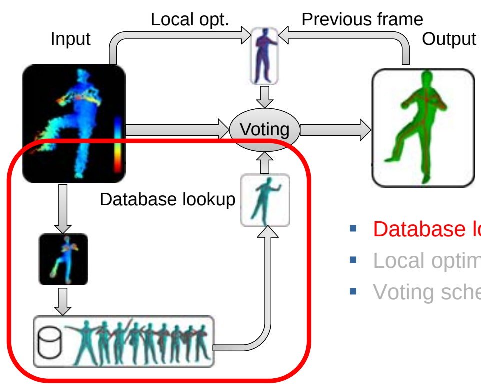
Input: Depth image Output: 3D pose



Data-Driven Motion Reconstruction

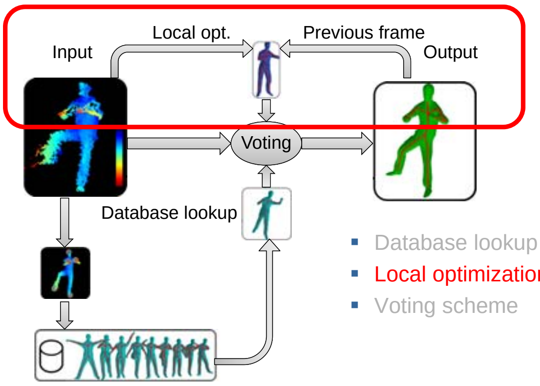


Data-Driven Motion Reconstruction



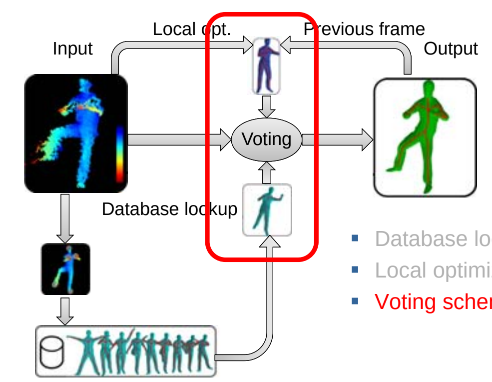
- Database lookup
- Local optimization
- Voting scheme

Data-Driven Motion Reconstruction



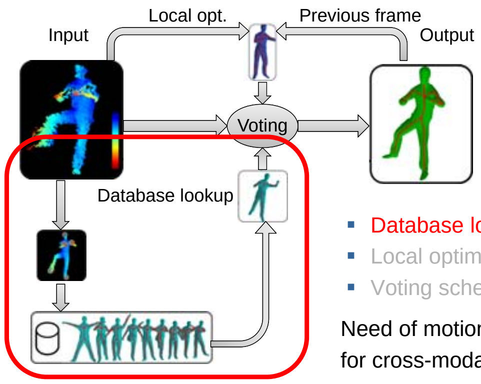
- Database lookup
- Local optimization
- Voting scheme

Data-Driven Motion Reconstruction



- Database lookup
- Local optimization
- Voting scheme

Database Lookup



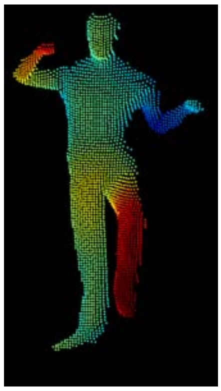
- Database lookup
- Local optimization
- Voting scheme

Need of motion features
for cross-modal comparison

Depth Image Features

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

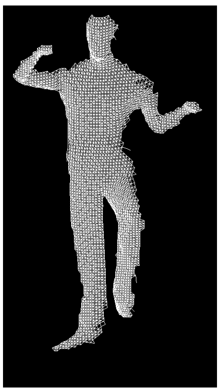
- Point cloud



Depth Image Features

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

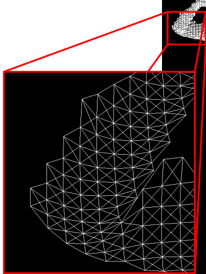
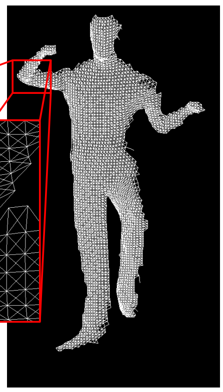
- Point cloud
- Graph



Depth Image Features

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

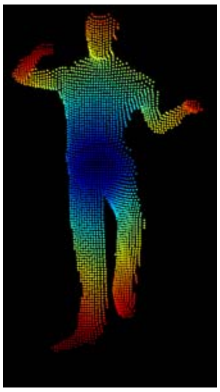
- Point cloud
- Graph

Depth Image Features

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

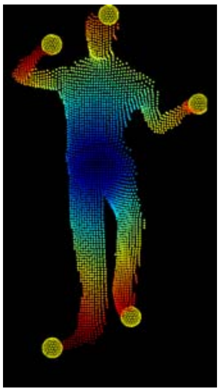
- Point cloud
- Graph
- Distances from root



Depth Image Features

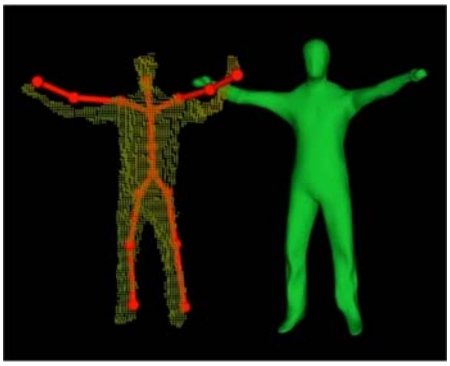
[Plagemann, Ganapathi, Koller, Thrun, ICRA 2010]

- Point cloud
- Graph
- Distances from root
- Geodesic extrema

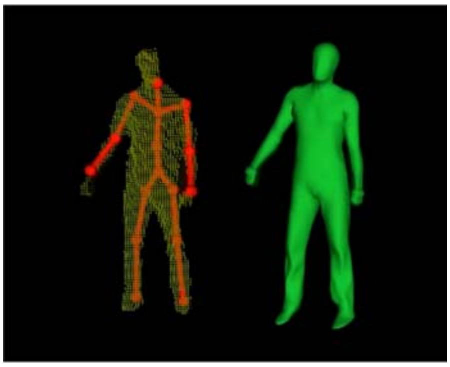


Observation: First five extrema often correspond to end-effectors and head

Database Lookup



Local Optimization

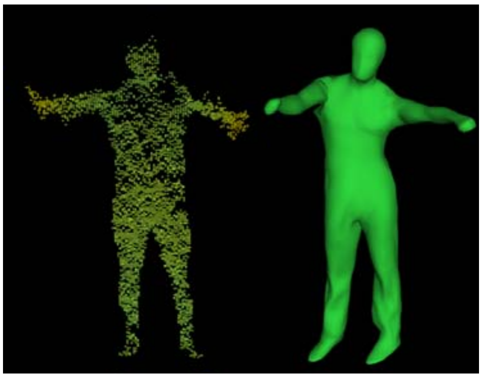


Voting Scheme

- 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

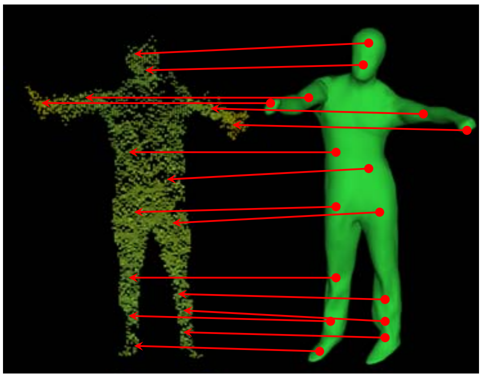
Voting Scheme

Distance measure



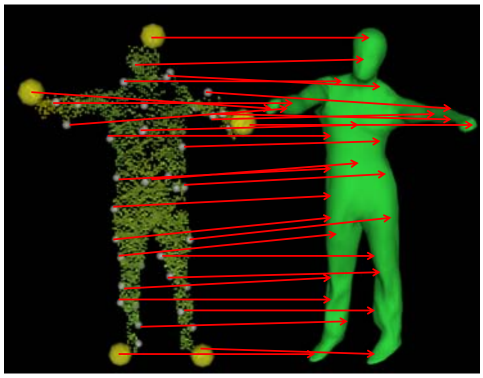
Voting Scheme

Distance measure (Hausdorff)

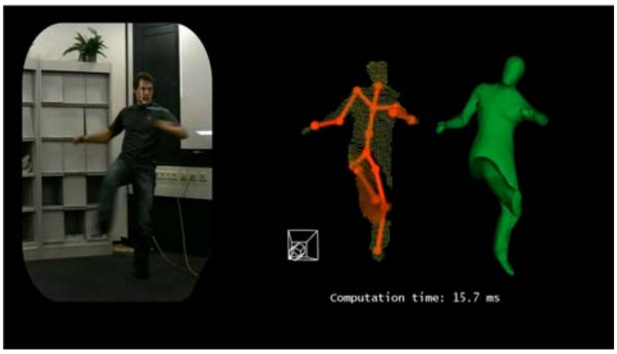


Voting Scheme

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

Informed Feature Representations

- Exploit model assumptions
  - Equal-tempered scale
  - Kinematic chain

Features with explicit meaning.

- Deal with variances on feature level
  - Enhancing timbre invariance
  - Relational features
  - Quantized motion templates

Makes subsequent steps more robust and efficient!

- Consider requirements for specific application
  - Explicit information often not required
  - Mid-level features

Avoid making problem harder as it is.

Conclusions

Selected Publications (Music Processing)

- M. Müller, P.W. Ellis, A. Klapuri, G. Richard (2011):
  Signal Processing for Music Analysis.
  IEEE Journal of Selected Topics in Signal Processing, Vol. 5, No. 6, pp. 1088-1110.
- P. Grosche and M. Müller (2011):
  Extracting Predominant Local Pulse Information from Music Recordings.
  IEEE Trans. on Audio, Speech & Language Processing, Vol. 19, No. 6, pp. 1688-1701.
- M. Müller, M. Clausen, V. Konz, S. Ewert, C. Fremerey (2010):
  A Multimodal Way of Experiencing and Exploring Music.
  Interdisciplinary Science Reviews (ISR), Vol. 35, No. 2.
- M. Müller and S. Ewert (2010):
  Towards Timbre-Invariant Audio Features for Harmony-Based Music.
  IEEE Trans. on Audio, Speech & Language Processing, Vol. 18, No. 3, pp. 649-662.
- F. Kurth, M. Müller (2008):
  Efficient Index-Based Audio Matching.
  IEEE Trans. Audio, Speech & Language Processing, Vol. 16, No. 2, 382-395.
- M. Müller (2007):
  Information Retrieval for Music and Motion.
  Monograph, Springer, 318 pages

Selected Publications (Motion Processing)

- J. Tautges, A. Zinke, B. Krüger, J. Baumann, A. Weber, T. Helten, M. Müller, H.-P. Seidel, B. Eberhardt (2011):
  Motion Reconstruction Using Sparse Accelerometer Data.
  ACM Transactions on Graphics (TOG) , Vol. 30, No. 3
- A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, C. Theobalt (2011):
  A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera.
  Proc. International Conference on Computer Vision (ICCV)
- G. Pons-Moll, A. Baak, T. Helten, M. Müller, H.-P. Seidel, B. Rosenhahn (2010):
  Multisensor-Fusion for 3D Full-Body Human Motion Capture.
  Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- A. Baak, B. Rosenhahn, M. Müller, H.-P. Seidel (2009):
  Stabilizing Motion Tracking Using Retrieved Motion Priors.
  Proc. International Conference on Computer Vision (ICCV)
- M. Müller, T. Röder, M. Clausen (2005):
  Efficient Content-Based Retrieval of Motion Capture Data.
  ACM Transactions on Graphics (TOG) , Vol. 24, No. 3, pp. 677-685, (SIGGRAPH)