Analysis and Retrieval Techniques for Motion and Music Data

Meinard Müller[†]

Saarland University and MPI Informatik, Campus E1 4, 66123 Saarbrücken Germany
`meinard@mpi-inf.mpg.de`

Abstract
In this tutorial, we study fundamental algorithms and concepts for the analysis, classification, indexing, and retrieval of time-dependent data streams considering motion capture data as well as waveform-based music data as examples. Important aspects concern the design of suitable features, the notion of similarity used to compare data streams, as well as data organization. One general goal of this tutorial is to highlight the interplay between modeling, experimentation, and mathematical theory as well as to give some insights into active research fields.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Systems]: Information Search and Retrieval—H.5.5 [Information Systems]: Sound and Music Computing—I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation
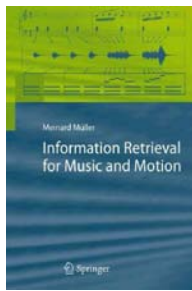
1. Introduction

Recent years have seen enormous advances in computerization and digitization as well as a corresponding growth in the use of information technology allowing users to access and experience multimedia content on an unprecedented scale. In this context, great efforts have been directed towards the development of techniques for searching and extracting useful information from huge amounts of stored data. In particular for textual information, powerful search engines have been implemented that provide efficient browsing and retrieval within billions of textual documents. For other types of multimedia data such as music, image, video, 3D shape, or 3D motion data, traditional retrieval strategies rely on textual annotations or metadata attached to the documents. Since the manual generation of descriptive labels is infeasible for large datasets, one needs fully automated procedures for data annotation as well as efficient *content-based* retrieval methods that only access the raw data itself without relying on the availability of annotations. A general retrieval scenario, which has attracted a large amount of attention in the field of multimedia information retrieval, is based on the *query-by-example paradigm*: given a query in form of a data fragment, the task is to automatically retrieve all documents from the database containing parts or aspects similar to the query. Here, the notion of similarity, which strongly depends on the respective application or on a person's perception, is of crucial importance in comparing the data. Frequently, multimedia objects, even though similar from a structural or semantic point of view, may reveal significant spatial or temporal differences. In other words, *semantic similarity* does not necessarily imply *numerical similarity*. This makes content-based multimedia retrieval a challenging research field with many yet unsolved problems.

In this tutorial, we introduce concepts and algorithms for robust and efficient information retrieval by means of two different types of multimedia data: human motion data and waveform-based music data. For both domains, motion and music, semantically related objects typically exhibit a large range of variations concerning temporal, spatial, spectral, or dynamic properties. We study fundamental strategies for handling object deformations and variability in the given data with a view to real-world retrieval and browsing applications. Here, one important principle, which is applicable to general multimedia data, is to already absorb variations that are to be left unconsidered in the searching process at the feature level. This strategy makes it possible to use relatively strict and efficient matching techniques.


In these notes, we provide motivating and domain-specific introductions of the information retrieval problems raised in this tutorial closely following the author's monograph titled *Information Retrieval for Music and Motion* [Mül07]. Section 2 covers the motion and Section 3 the music domain. In each of these sections, we also give pointers to the literature relevant to motion and music retrieval, respectively, which may serve the reader as entry points for further research. Finally, in Section 4, we discuss some general concepts for content-based information retrieval, which apply to both domains, music and motion, and beyond. For a comprehensive account on the problems and concepts raised in this tutorial, we refer to the book [Mül07].

2. Motion Retrieval

Generally speaking, motion capturing is referred to as the process of recording real moving objects and creating an abstract three-dimensional digital representation of these motions. The resulting *motion capture data* or simply *mocap data* constitutes the basis for applications in various fields such as gait analysis, rehabilitation, physical therapy, biomechanical research, animal science, and sports performance analysis. With increasing importance, mocap data has also been used in computer animation to create realistic motions for both movies and video games. Here, one typically proceeds in several steps: after planning the motion capture shoots, the motions are performed by live actors and captured by the mocap system. The resulting data is cleaned and suitably processed using editing and blending techniques, and finally the motions are mapped to the animated characters.

Present motion capture systems can track and record human motions at high spatial and temporal resolutions, which makes it possible to capture even subtle movements and nuances of motion patterns. This is the reason why mocap-based animations usually look very realistic and natural compared with traditional hand-animated 3D models, where it is too time consuming and difficult to accurately represent motion details. On the downside, the lifecycle of a motion clip in an animation production is very short. Typically, a motion clip is captured, incorporated in a single 3D scene, and then never used again. Furthermore, the adjustment of once captured motion data to fit certain constraints and user-specified needs is a difficult problem. Therefore, it often seems easier to record new data rather than trying to manipulate the data. However, motion capturing is an expensive process—high-quality mocap systems easily cost more than one hundred thousand dollars including digital video cameras and software. For efficiency and cost reasons, the reuse of mocap data as well as methods for modifying and adapt-
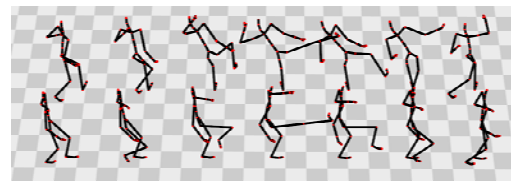
Figure 1: *Top: 7 poses from a side kick sequence.* **Bottom:** *corresponding poses for a front kick. The two kicking motions, even though being similar in some semantic sense, exhibit significant spatial and temporal differences.*

ing existing motion clips are gaining in importance. Here, an active field of research is the application of editing, morphing, and blending techniques for the creation of new, realistic motions from prerecorded motion clips. Such techniques depend on motion capture databases covering a broad spectrum of motions in various characteristics. Only in the last few years, larger collections of motion material have become publicly available.

Prior to reusing and processing motion capture material, one has to solve the fundamental problem of *identifying* and *extracting* suitable motion clips scattered in a given database. Traditional approaches to motion retrieval rely on manually generated annotations, where the motions to be identified are roughly described in words such as "a kick of the right foot followed by a punch". Retrieval is then performed at the metadata level. Since the manual generation of descriptive labels is infeasible for large datasets, one needs efficient *content-based* retrieval methods that only access the raw data itself. Here, a typical query mode is based on the *query-by-example paradigm*, which has attracted a large amount of attention in the field of information retrieval: given a query in form of a data fragment, the task is to automatically retrieve all documents from the database containing parts or aspects similar to the query.

The crucial point in content-based motion retrieval is the notion of *similarity* used to compare different motions. Typically, the variations may concern the spatial as well as the temporal domain. For example, the kicks shown in Figure 1 describe the same kind of motion even though they differ considerably with respect to motion speed as well as the direction and height of the kick. Due to these spatio-temporal motion variations, content-based retrieval of motion capture data constitutes a difficult and time-consuming problem. Recent approaches to motion retrieval apply techniques such as dynamic time warping, which, however, are not applicable to large datasets due to their quadratic space and time complexity. In view of more efficient and robust motion retrieval strategies, the following observation is of fundamental importance: opposed to other data formats such as image or video, motion capture data is based on a kinematic chain, which represents the human skeleton. This underlying model can be exploited by looking for geometric rela-

tions between specified body points of a pose, where the relations possess an explicit semantic meaning. For example, suppose the user is looking for all right foot kicking motions contained in the database. Even though there may be large variations between different kicking motions, all such motions share some common characteristics: first the right knee is stretched, then bent, and finally stretched again, while the right foot is raised during this process. Afterwards, the right knee is again bent and then stretched, while the right foot drops to the floor again. In other words, by only considering the two simple boolean relations "right knee bent or not" and "right foot raised or not" in the temporal context, one can exclude all motions in the database that do not reveal the characteristic progression of relations as described above. With this strategy, one can typically cut down the search space very efficiently from several hours to a couple of minutes of motion capture data, which can then by analyzed and processed by more refined techniques.

Links to the Literature

We now give some references, which may serve the reader as entry points to the literature on motion synthesis, analysis, and retrieval. The bibliography is not meant to be comprehensive. Further links to the literature can be found in the cited papers.

In view of massively growing multimedia databases of various types and formats, efficient methods for indexing and content-based retrieval have become an important issue. Vast literature exists on content-based database retrieval including text [WMB99], image, and video data [BHL*03], as well as 3D models [FMK*03]. For the music scenario, Clausen and Kurth [CK04] give a unified approach to content-based retrieval; their group theoretical concepts generalize to other domains as well. The problem of indexing large time series databases has also attracted great interest in the database community, see, e.g., Keogh [Keo02] and Last et al. [LKB04] and the references therein.

The reuse of motion capture data via editing and morphing techniques has been a central topic in data-driven computer animation for more than a decade, starting with [BW95, WP95]. Since then, many different methods have been suggested to create new, realistic motions from prerecorded motions; see, for example, [AFO03, CH05, CHP07, GP00, KG03, KG04, PB02, ZMCF05, ZMM*07] and the references therein. Only recently, motion capture data has become publicly available on a larger scale (e.g., CMU [CMU03] or HDM05 [MRC*05]), reinforcing the demand for efficient indexing and retrieval methods. Because of possible spatio-temporal variations, the difficult task of identifying similar motion segments still bears open problems. Most of the approaches to motion comparison are based on features that are semantically close to the raw data, using 3D positions, 3D point clouds, joint angle representations, or PCA-reduced versions thereof, see [FF05, HPP05,

KPZ*04, KG04, SKK04, WCYL03]. One problem of such features is their sensitivity towards pose deformations, as may occur in semantically related motions. To achieve more robustness, Müller et al. [MRC05] introduced relational features and adaptive temporal segmentation, absorbing spatio-temporal variations already at the feature level. A similar strategy has been used by Liu et al. [LZWM05], who transform motions into sequences of cluster centroids, which absorb spatio-temporal variations. Motion comparison is then performed on these sequences. Based on relational motion features, Baak et al. [BMS08] introduced a keyframe-based motion search algorithm, which allows for explicitly controlling the degree of admissible deformations between the queried keyframes, while being efficient using an inverted file index.

Automatic motion annotation and classification are closely related to the retrieval problem and are important tasks in view of motion reuse. Arikan and Forsyth [AFO03] propose a semi-automatic annotation procedure for motion data using support vector machine classifiers. Ramanan and Forsyth [RF03] apply this annotation technique for 3D motion data as a preprocessing step for the automatic annotation of 2D video recordings of human motion, using HMMs to match the 2D data with the 3D data. Rose et al. [RCB98] group similar example motions into "verb" classes to synthesize new, user-controlled motions by suitable interpolation techniques. Müller et al. [MR06] introduced the concept of motion templates, which allows for grasping the essence of an entire class of motions within an explicit matrix representation. Several approaches for automatic classification and recognition of motion patterns are based on Hidden Markov models (HMMs), which are also a flexible tool to capture spatio-temporal variations, see, e. g., [BH00, GG04, WB99]. Temporal segmentation of motion data can be viewed as another form of annotation, where consecutive, semantically related frames are organized into groups, see, e. g., [BSP*04, FMJ02].

For an overview of various techniques that are useful for organizing, processing, and navigating mocap databases, we refer to the course notes [FHP07]. Many useful links to the motion synthesis and tracking literature can be found in the review [FAI*06], which is intended to cross-fertilize ideas about motion representation between the animation and computer vision communities. Further references as well as a detailed account on motion representations and similarity aspects can be found in the book [Mül07].

3. Music Retrieval

For music, there is a vast amount of digitized data as well as a variety of associated data representations, which describe music at various semantic levels. Typically, digital music collections contain a large number of relevant digital documents for one musical work, which are given in various digital formats and in multiple realizations. For example,
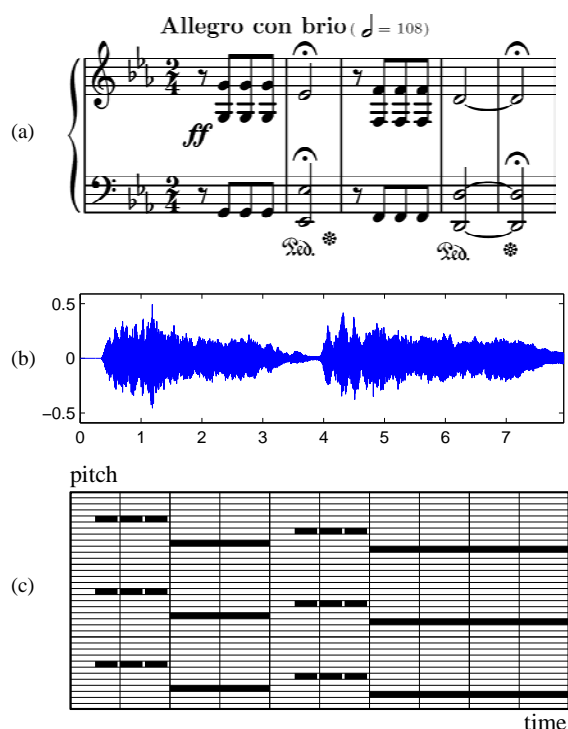
Figure 2: *(a) Score representation of the first five measures of Beethoven's Fifth Symphony in a piano reduction (from [Mutopia Project06]). (b) Waveform of a Bernstein interpretation of these measures. (c) Piano roll representation. The black rectangles correspond to the notes shown in (a).*

in the case of Beethoven's Fifth Symphony, a digital music library may contain the scanned pages of some particular score edition. Or the score may be given in a digital music notation file format, which encodes the page layout of sheet music in a machine-readable form. Furthermore, the library may contain various CD recordings such as the interpretations by Karajan and Bernstein, some historical recordings by Furthwängler and Toscanini, Liszt's piano transcription of Beethoven's Fifth played by Glenn Gould, as well as a synthesized version of a corresponding MIDI file. The different interpretations of Beethoven's Fifth often reveal large variations regarding tempo, dynamics, articulation, tuning, or instrumentation.

As illustrated by the Beethoven example, there are various digital manifestations of a musical work differing in format and content, see Figure 2. In the field of *music information retrieval* (MIR), great efforts have been directed towards the development of technologies that allow users to access and explore music in all its different facets. For example, during playback of some CD recording, a digital music player of the future presents the corresponding musical score while highlighting the current playback position within the score. On

demand, additional information about melodic and harmonic progression or rhythm and tempo is automatically presented to the listener. A suitable user interface displays the musical structure of the current piece of music and allows the user to directly jump to any key part within the recording without tedious fast-forwarding and rewinding. Furthermore, the listener is equipped with a Google-like search engine that enables him to explore the entire music collection in various ways: the user creates a query by specifying a certain note constellation or some harmonic or rhythmic pattern by whistling a melody or simply by selecting a short passage from a CD recording; the system then provides the user with a ranked list of available music excerpts from the collection that are musically related to the query. For example, querying a twenty-second excerpt of a Bernstein interpretation of the theme of Beethoven's Fifth, the system will return all other corresponding music clips in the database. This includes the repetition of the theme in the exposition or in the recapitulation within the same interpretation as well as the corresponding excerpts in all recordings of the same piece interpreted by other conductors. An advanced search engine is also capable of automatically identifying the theme even in the presence of significant variations, thus handling arrangements such as Liszt's piano transcription, synthesized versions, or rhythmically accompanied pop versions of Beethoven's Fifth.

Even though significant progress has been made in the development of advanced music players, there are still many yet unsolved problems in content-based music browsing and retrieval, which are due to the heterogeneity and complexity of music data. Here, *content-based* means that in the comparison of music data, the system only makes use of the raw data itself, rather than relying on manually generated metadata such as keywords or other symbolic descriptions. While text-based retrieval of music documents using the composer's name, the opus number, or lyrics can be handled by means of classical database techniques, purely content-based music retrieval constitutes a difficult research problem. How should a retrieval system be designed, if the user's query consists of a whistled melody fragment or a short excerpt of some CD recording? How can (symbolic) score data be compared with the content of (waveform-based) CD recordings? What are suitable notions of similarity that capture certain (user-specified) musical aspects while disregarding admissible variations concerning, e. g., the instrumentation or articulation? How can the musical structure, reflected by repetitive and musically related patterns, be automatically derived from a CD recording? These questions only reflect a small fraction of current MIR research topics that are closely related to automatic music analysis.

Links to the Literature

In the last decade, music information retrieval (MIR) has become an active and multidisciplinary research field. Cen-

tral MIR problems concern music information handling and retrieval, automated music recognition and classification, the design and extraction of musically relevant audio features, or the development of novel user interfaces. Because of the diversity and richness of music, MIR research brings together experts from a multitude of research fields ranging from information science, audio engineering, computer science, musicology, music theory, library science, to law and business. For a general account on the multidisciplinary and multifaceted challenges of MIR research, we refer to the overview article by Downie [Dow03]. Recently, Pardo [Par06] has edited a series of short overview articles, which outline current MIR research problems from a more technical point of view. The tutorial by Orio [Ori06] summarizes fundamental issues on music representations, user interaction, music processing, and specifications of MIR systems. A detailed account on various music analysis problems from a multidisciplinary viewpoint, including aspects from psychoacoustics and music perception, can be found in the Ph. D. thesis by Scheirer [Sch00]. The book edited by Klapuri and Davy [KD06] deals with the central MIR problem of automatic *music transcription* comprising prominent subproblems such as rhythm analysis, fundamental frequency analysis, source separation, and musical instrument classification. The book also covers important signal processing and pattern recognition methods from a statistical point of view. The book by Müller [Mül07] gives an account on state-of-the-art techniques for recent MIR problems including *music synchronization*, *audio matching*, and *audio structure analysis*. It also covers fundamentals on music representations, digital signal processing, feature design, and dynamic time warping. The comprehensive book by Mazzola [Maz02] deals with mathematical music theory and introduces a conceptual basis for music composition, analysis, and performance. Even though this book presumes a thorough understanding of deep mathematical concepts, it gives a unified approach to a large number of musical aspects and constitutes a rich source of ideas worth considering in MIR research. Finally, we refer to the annual International Conference on Music Information Retrieval (ISMIR), which constitutes a multidisciplinary platform for researchers involved in work on accessing digital musical materials. The proceedings of this conference, which are available online and can be accessed via the conference homepage [ISMIR], contain a broad spectrum of research papers and reflect the state of the art in MIR research.

4. Methods and Strategies

In this tutorial, we introduce concepts and algorithms for robust and efficient multimedia retrieval in the presence of variations using 3D motion capture data and waveform-based music data as our main examples. Even though these two types of multimedia data bear no immediate semantic relationship, music and motion share—from an abstract point of view—some common properties. First of all, both kinds

of data are subject to relatively strong model assumptions. For example, motion capture data is based on an explicit model in the form of a kinematic chain, which represents the human skeleton. This underlying model can be exploited by looking for geometric relations between specific body points of a pose, where the relations possess an explicit semantic meaning. For example, using the two boolean relations that check whether the "right knee is bent or not" and whether the "right foot is raised or not" one can capture important characteristics of a right foot kicking motion. Similarly, most Western music is based on the traditional equal-tempered chromatic scale implying the occurrences of distinguished frequency components in an audio recording. This assumption can be exploited by extracting, e. g., pitch-based audio features that allow for some direct, musically relevant interpretation. As a first important topic of this tutorial, we discuss feature classes that closely correlate to semantic aspects of the underlying data while showing a high degree of invariance to irrelevant deformations and variations.

A further common property of motion and music data is their temporal dimension. In both cases the data can be transformed into time-dependent feature sequences that reflect the changing characteristics in the raw data over time. For example, a combination of simple boolean relations in the temporal context is often sufficient to characterize specific classes of similar motions. Similarly, an audio recording can be transformed into a sequence of pitch-based feature vectors that closely relate to the harmonic progression of the underlying piece of music. As a second important topic, we explain how the temporal order of the extracted features can be exploited to identify similar motion and music fragments.

Note that for both domains, motion and music, semantically related objects exhibit a large range of variations concerning temporal, spatial, spectral, or dynamic properties. To handle deformations and variability in the objects, we show how to simultaneously employ various invariance and fault-tolerance mechanisms at different conceptual levels. First, by employing deformation-tolerant features, one can already absorb a high degree of the undesired variations at the feature level. Second, to compare features or short feature sequences, we introduce enhanced local cost measures that are suited to handle local temporal and spatial variations. Third, by using global similarity measures that are based on mismatch, fuzzy, and time-warping concepts, we add another degree of robustness and fault-tolerance at a more global level.

In addition to deformation-tolerance and robustness, the question of efficiency is of fundamental importance in content-based multimedia retrieval, in particular in view of large amounts of data that have to be processed and searched. As a further topic of this tutorial, we discuss various methods for speeding up computations, including data reduction and clustering techniques, multiscale (coarse-to-fine) and preselection strategies, and index-based search and retrieval pro-

cedures. As it turns out, an overall procedure for content-based multimedia retrieval typically suffers from the trade-off between the capability of handling object deformations on the one hand and retrieval efficiency on the other hand. For example, time-warping strategies are powerful in coping with temporal deformations but are generally expensive with respect to computation time and memory. In contrast, index-based strategies may afford efficient data retrieval that scales to large datasets, however, at the expense of being rather inflexible to variations. Here, the usage of fault-tolerance mechanisms such as mismatch or fuzzy search restores flexibility to some extent, but at the price of increased computational cost. We discuss how a particular choice or combination of methods may yield efficient retrieval procedures suitable in use for real-world application scenarios.

5. Intended and Expected Audience

In this tutorial, we discuss a number of current research problems in multimedia information retrieval covering aspects from information science, computer animation, and digital signal processing. To account for an interdisciplinary audience, we provide the necessary background information and give numerous motivating examples from the music and motion domain. By providing illustrative examples and by working with pictures (rather than with formulas), this tutorial is particularly designed for non-experts and researchers from various fields in academia and industry.

6. Biography of the Presenter

Meinard Müller studied mathematics and computer science at Bonn University, Germany, where he received both a Master's degree in mathematics and his PhD in computer science in the year 1997 and 2001, respectively. In 2002/2003, he conducted postdoctoral research in combinatorics at the Mathematical Department of Keio University, Japan. In 2007, he obtained his Habilitation in the field of multimedia retrieval. His Habilitation thesis appeared as Springer monograph titled *Information Retrieval for Music and Motion* [Mül07]. Currently, Meinard Mülller is a member of the Saarland University and the Max-Planck-Institut für Informatik working as a senior researcher within the Cluster of Excellence *Multimodal Computing and Interaction.* His recent research interests include content-based multimedia retrieval, audio signal processing, music processing, as well as the analysis and classification of 3D motion capture data.

References

[AFO03] ARIKAN O., FORSYTH D. A., O'BRIEN J. F.: Motion synthesis from annotations. *ACM Trans. Graph. 22*, 3 (2003), 402–408.

[BH00] BRAND M., HERTZMANN A.: Style machines. In *Proc. ACM SIGGRAPH '00* (2000), Computer Graphics Proc., ACM Press, pp. 183–192.

[BHL*03] BAKKER E. M., HUANG T. S., LEW M. S., SEBE N., ZHOU X. S. (Eds.):. *Proc. 2nd Intl. Conf. Image and Video Retrieval, CIVR 2003, Urbana-Champaign, IL, USA* (2003), vol. 2728 of *LNCS*, Springer.

[BMS08] BAAK A., MÜLLER M., SEIDEL H.-P.: An efficient algorithm for keyframe-based motion retrieval in the presence of temporal deformations. In *Proc. ACM International Conference on Multimedia Information Retrieval, Vancouver, Canada* (2008).

[BSP*04] BARBIC J., SAFONOVA A., PAN J.-Y., FALOUTSOS C., HODGINS J. K., POLLARD N. S.: Segmenting motion capture data into distinct behaviors. In *GI '04: Proc. Graphics interface* (2004), Canadian Human-Computer Communications Society, pp. 185–194.

[BW95] BRUDERLIN A., WILLIAMS L.: Motion signal processing. In *Proc. ACM SIGGRAPH 95* (1995), Computer Graphics Proc., ACM Press, pp. 97–104.

[CH05] CHAI J., HODGINS J. K.: Performance animation from low-dimensional control signals. *ACM Trans. Graph. 24*, 3 (2005), 686–696.

[CHP07] COOPER S., HERTZMANN A., POPOVIĆ Z.: Active learning for real-time motion controllers. *ACM Trans. Graph. 26*, 3 (2007), 5.

[CK04] CLAUSEN M., KURTH F.: A unified approach to content-based and fault tolerant music recognition. *IEEE Transactions on Multimedia 6*, 5 (2004), 717–731.

[CMU03] CMU: Carnegie-Mellon Mocap Database. http://mocap.cs.cmu.edu, 2003.

[Dow03] DOWNIE J. S.: Music information retrieval. *Annual Review of Information Science and Technology (Chapter 7) 37* (2003), 295–340.

[FAI*06] FORSYTH D. A., ARIKAN O., IKEMOTO L., O'BRIEN J., RAMANAN D.: Computational studies of human motion: part 1, tracking and motion synthesis. *Found. Trends. Comput. Graph. Vis. 1*, 2 (2006), 77–254.

[FF05] FORBES K., FIUME E.: An efficient search algorithm for motion data using weighted PCA. In *Proc. 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2005), ACM Press, pp. 67–76.

[FHP07] FALOUTSOS C., HODGINS J., POLLARD N.: Database techniques with motion capture. In *SIGGRAPH '07: ACM SIGGRAPH 2007 courses* (New York, NY, USA, 2007), ACM, p. 1.

[FMJ02] FOD A., MATARIC M. J., JENKINS O. C.: Automated derivation of primitives for movement classification. *Auton. Robots 12*, 1 (2002), 39–54.

[FMK*03] FUNKHOUSER T., MIN P., KAZHDAN M., CHEN J., HALDERMAN A., DOBKIN D., JACOBS D.: A search engine for 3D models. *ACM Trans. Graph. 22*, 1 (2003), 83–105.

[GG04] GREEN R., GUAN L.: Quantifying and recognizing human movement patterns from monocular video images: Part I. *IEEE Transactions on Circuits and Systems for Video Technology 14*, 2 (February 2004), 179–190.

[GP00] GIESE M., POGGIO T.: Morphable models for the analysis and synthesis of complex motion patterns. *IJCV 38*, 1 (2000), 59–73.

[HPP05] HSU E., PULLI K., POPOVIĆ J.: Style translation for human motion. *ACM Trans. Graph. 24*, 3 (2005), 1082–1089.

[ISMIR] ISMIR: Hompage of the international conferences on music information retrieval. http://www.ismir.net.

[KD06] KLAPURI A., DAVY M. (Eds.): *Signal Processing Methods for Music Transcription*. Springer, New York, 2006.

[Keo02] KEOGH E.: Exact indexing of dynamic time warping. In *Proc. 28th VLDB Conf., Hong Kong* (2002), pp. 406–417.

[KG03] KOVAR L., GLEICHER M.: Flexible automatic motion blending with registration curves. In *Proc. 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2003), Eurographics Association, pp. 214–224.

[KG04] KOVAR L., GLEICHER M.: Automated extraction and parameterization of motions in large data sets. *ACM Trans. Graph. 23*, 3 (2004), 559–568.

[KPZ*04] KEOGH E. J., PALPANAS T., ZORDAN V. B., GUNOPULOS D., CARDLE M.: Indexing large human-motion databases. In *Proc. 30th VLDB Conf., Toronto* (2004), pp. 780–791.

[LKB04] LAST M., KANDEL A., BUNKE H. (Eds.): *Data Mining In Time Series Databases*. World Scientific, 2004.

[LZWM05] LIU G., ZHANG J., WANG W., MCMILLAN L.: A system for analyzing and indexing human-motion databases. In *Proc. 2005 ACM SIGMOD Intl. Conf. on Management of Data* (2005), ACM Press, pp. 924–926.

[Maz02] MAZZOLA G.: *The topos of music*. Birkhäuser, 2002.

[MR06] MÜLLER M., RÖDER T.: Motion templates for automatic classification and retrieval of motion capture data. In *SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2006), ACM Press, pp. 137–146.

[MRC05] MÜLLER M., RÖDER T., CLAUSEN M.: Efficient content-based retrieval of motion capture data. *ACM Trans. Graph. 24*, 3 (2005), 677–685.

[MRC*05] MÜLLER M., RÖDER T., CLAUSEN M., EBERHARDT B., KRÜGER B., WEBER A.: *Documentation: Mocap Database HDM05*. Computer Graphics Technical Report CG-2007-2, Universität Bonn, June 2007 http://www.mpi-inf.mpg.de/resources/HDM05.

[Mül07] MÜLLER M.: *Information Retrieval for Music and Motion*. Springer, 2007.

[Mutopia Project06] MUTOPIA PROJECT: Music free to download, print out, perform and distribute. http://www.mutopiaproject.org, 2006.

[Ori06] ORIO N.: Music retrieval: A tutorial and review. *Foundation and Trends in Information Retrieval 1*, 1 (Nov. 2006), 1–90.

[Par06] PARDO B.: Music information retrieval. *Special Issue, Commun. ACM 49*, 8 (2006), 28–58.

[PB02] PULLEN K., BREGLER C.: Motion capture assisted animation: Texturing and synthesis. *ACM Trans. Graph.* (2002), 501–508.

[RCB98] ROSE C., COHEN M. F., BODENHEIMER B.: Verbs and adverbs: Multidimensional motion interpolation. *IEEE Comput. Graph. Appl. 18*, 5 (1998), 32–40.

[RF03] RAMANAN D., FORSYTH D. A.: Automatic annotation of everyday movements. In *Advances in Neural Information Processing Systems 16* (2003).

[Sch00] SCHEIRER E. D.: *Music-Listening Systems*. PhD thesis, Program in Media Arts and Sciences, MIT, 2000.

[SKK04] SAKAMOTO Y., KURIYAMA S., KANEKO T.: Motion map: image-based retrieval and segmentation of motion data. In *Proc. 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2004), ACM Press, pp. 259–266.

[WB99] WILSON A. D., BOBICK A. F.: Parametric hidden markov models for gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 9 (1999), 884–900.

[WCYL03] WU M.-Y., CHAO S., YANG S., LIN H.: Content-based retrieval for human motion data. In *16th IPPR Conf. on Computer Vision, Graphics and Image Processing* (2003), pp. 605–612.

[WMB99] WITTEN I. H., MOFFAT A., BELL T. C.: *Managing Gigabytes*. Morgan Kaufmann Publishers, 1999.

[WP95] WITKIN A., POPOVIĆ Z.: Motion warping. In *Proc. ACM SIGGRAPH 95* (1995), Computer Graphics Proc., ACM Press/ACM SIGGRAPH, pp. 105–108.

[ZMCF05] ZORDAN V. B., MAJKOWSKA A., CHIU B., FAST M.: Dynamic response for motion capture animation. *ACM Trans. Graph. 24*, 3 (2005), 697–701.

[ZMM*07] ZORDAN V., MACCHIETTO A., MEDIN J., SORIANO M., WU C.-C., METOYER R., ROSE R.: Anticipation from example. In *VRST '07: Proceedings of the 2007 ACM symposium on Virtual reality software and technology* (New York, NY, USA, 2007), ACM, pp. 81–84.