8

Enhancing Multimedia Search Using Human Motion

Kevin Adistambha[1], Stephen Davis[1], Christian Ritz[1],
Ian S. Burnett[2] and David Stirling[1]
[1]*University of Wollongong,*
[2]*RMIT University,*
Australia

1. Introduction

Over the last few years, there has been an increase in the number of multimedia-enabled devices (e.g. cameras, smartphones, etc.) and that has led to a vast quantity of multimedia content being shared on the Internet. For example, in 2010 thirteen million hours of video uploaded to YouTube (http://www.youtube.com). To usefully navigate this vast amount of information, users currently rely on search engines, social networks and dedicated multimedia websites (such as YouTube) to find relevant content. Efficient search of large collections of multimedia requires metadata that is human-meaningful, but currently multimedia sites generally utilize metadata derived from user-entered tags and descriptions. These are often vague, ambiguous or left blank, which makes search for video content unreliable or misleading. Furthermore, a large majority of videos contain people, and consequently, human movement, which is often not described in the user entered metadata.

To compensate for the lack of metadata, which is crucial for efficient search, there has been research into automated techniques for the extraction of metadata to create multimedia description formats. The latter can generally be described as *temporal* (e.g. a point in time within a video when something happens), or *non-temporal* (e.g. a video was created by John Doe). Popular metadata formats such as Dublin Core (Weibel et al. 1998) are typically used as a *non-temporal* description format. e.g., using Dublin Core to describe the whole video clip, while other description formats, such as MPEG-7 (Martínez 2004), are capable of describing multimedia in both a *temporal* (i.e., Color Structure Descriptor) and *non-temporal* (i.e., parts of MPEG-7 Multimedia Description Schemes (MDS)) manner. In the context of search, both temporal and non-temporal metadata sets are both important.

One approach to multimedia search for action sequences is description of the action in terms of a *Subject-Verb-Object* construct similar to RDF (W3C 2004). For example, one can search for "*John Doe runs and then jumps into a lake*", or "*Jane Doe runs in a circle by a tree*". In this analogy, Dublin Core might be used to provide the *Subject* description, and MPEG-7 the *Object* description: Dublin Core can describe a person in general terms (subject), and MPEG-7 can accurately describe a lake (object). However, using this construct, the crucial

description of the *Verb* is missing. The question becomes, how do we describe a person running, jumping, walking, etc.?

Consider a simple scenario where John Doe has just come back from a holiday and uploaded 100 videos taken while away all at once to YouTube. As a result of the single upload, John did not have the time to annotate all of his videos individually. When one of his friends wanted to search for the video that contained the funny moment that everyone was talking about (when he jumped into a lake), it was not possible to search for that exact moment using current description technologies. Instead, John's friends would need to look at each of the 100 videos to find it and it is likely that their enthusiasm would wane. Even if some of the metadata was automatically extracted/generated (Dublin Core: uploaded by John Doe, MPEG-7: spatio-temporal color descriptors), it is not currently possible to perform a human motion search in a temporal manner. With 35 hours of video uploaded every minute to YouTube (http://www.youtube.com), this is becoming a real, immediate problem to be solved. This chapter provides a solution to the missing verb description in the Subject-Verb-Object multimedia search for the case of human motion, by proposing a searchable human motion description that can be used in conjunction with existing metadata technologies

The rest of this chapter is structured as follows: Section 2 provides related work in human motion description area. Section 3 provides the detail of the new human motion description format that is simple and searchable. Section 4 concludes the chapter and provides possible future work in this area.

2. Related work

2.1 Multimedia description schemes

Multimedia description schemes describe content (e.g. video and motion capture) with different detail granularities in terms of their semantic description of the underlying raw data. Existing methods include the Dublin Core metadata that describes the media using high-level descriptors (e.g. title, subject, creator, creation date), and MPEG-7 as an overarching description format. MPEG-7 combines many description schemes that can describe the content of audiovisual data in terms of many aspects including color and sound characteristics. MPEG-7 Part 5 Multimedia Description Scheme (MDS) also provides a high level semantic description similar to Dublin Core (although Dublin Core was designed to describe formats not limited to multimedia). In the case of the *Subject-Verb-Object* analogy described in Section 1, the *Subject* and *Object* portion of multimedia description are adequately represented by both Dublin Core and MPEG-7, while the *Verb* portion is not. Unfortunately, these description schemes do not provide a temporal representation of human motion.

2.2 Motion capture based human motion description

With the advent of motion capture technologies and the relatively low cost of processing power nowadays, people have been recording an increasing amount of motion for games development and movies, and hence the problem of how to catalogue the motion and to search the resulting motion capture databases is becoming increasingly relevant. The state of the art in this area is the work of Mueller et. al. (Mueller et al. 2009) and Guerra-Filho et. al.

(Gutemberg Guerra-Filho & Yiannis Aloimonos 2007). Mueller developed the concept of a "motion template", where a number of example sequence of a given motion are analyzed and their common features recognized. The result of this analysis is the template for that specific motion. Some tolerance was designed into the system, so that minor differences in movements would still be recognized as the same motion (hence the term "template"). In contrast, Guerra-Filho developed the Human Action Language that takes inspiration from written language and defines a motion as a series of small actions connected together to form larger, more complex motions. The key concepts of the Human Action Language are compactness of description (describing an action with the least amount of symbols, called atoms), view-invariance (a 3D motion should be able to be projected into a 2D plane), selectivity (the ability to differentiate between different atoms) and reconstructivity (reconstruction of a described motion back into its motion capture representation). Barbic et al (Barbič et al. 2004) explores the use of Principal Component Analysis (PCA) and Gaussian Mixture Model (GMM) to perform automatic segmentation of motion capture data, due to their observation that a motion capture session tends to get longer as more motion is captured, especially if natural behavior of the actor during the capture session is desired. Barbic et. al. achieved high accuracy using the PCA approach. Similarly, Li et. al. (Li et al. 2007) explored the use of Singular Value Decomposition (SVD) to extract features in a continuous motion capture data to perform automatic motion segmentation into distinct motion classes. Most recently, VideoMocap by Wei et. al. (Wei & Chai 2010) provides a method to extract a motion capture-like 3D representation of the human body from 2D video. This is achieved by manually annotating the joints in keyframes in the video, and by using physics-based interpolation method to reconstruct the 3D skeleton of the human body accordingly. The work by Gu et. al.(Gu et al. 2010) further performed action and gait recognition on a reconstructed 3D model of the human body from video using a Hidden Markov Model (HMM) based approach; the reconstruction of the 3D human body by Gu et.al. combines manual annotation of the body joints with a hierarchical skeleton model similar to motion capture representation of the human body.

Although significant progress has been made in the area of human motion from a numerical point of view using GMM, PCA, SVD, HMM, or physics-based methods, annotating motion data in a structured manner (from either video or motion capture) for search purposes is noticeably absent. Advances in the form of Human Action Language and Motion Templates are bridging the gap between semantic and numeric descriptions of human motion, but both technologies are not designed with general multimedia search that can work together with existing metadata technologies.

2.3 Dance notation based human motion description

Labanotation (Laban & Lange 1975) was developed by Rudolf Laban to record dance movements using a series of symbols written on a special staff where each place in the staff denotes a major limb such as both legs, both arms, the torso, and the head. This popular notation served as the basis for some technologies for human motion descriptions.

The work by Nakamura (Nakamura & Hachimura 2006) provides a crossover between Labanotation and motion capture. The goal of Nakamura is to provide a well-formed Labanotation description of motion capture data in XML, called LabanXML and is defined using Document Type Definition (DTD). The work in LabanXML is further extended to include the full Labanotation as described in (Ann Hutchinson 1970) in the form of

MovementXML by Hatol (Hatol 2006). In contrast to the DTD approach used in LabanXML, MovementXML utilizes the more modern XML Schema approach while also providing a more complete Labanotation description. Both LabanXML and MovementXML provide direct translation from handwritten Labanotation directly into XML. Further, the work by Loke et al (Loke et al. 2005) explores the possibilities of using a Labanotation-based effort metric to serve as an input for a human-computer interface.

In terms of computer-based multimedia search, dance notation based techniques rely on musical concepts of "beats" and "measures" to provide timing information, and encapsulate the motion descriptions in the context of those timing parameters. These metrics do not align with the actual timing information in a recorded motion (e.g. in a video) that is based on the number of frames captured per second. Since arguably music-based timing information also depends on the type of motion being performed (e.g. a slow walk would have a slower "beat" compared to a sprint), objective and consistent timing information is required for a searchable human motion description.

2.4 Summary

Most of the literature describing human motion has focused on translating the raw data into low-level metadata, often at the frame or near-frame level. These techniques do not provide high level metadata that can describe human motion (e.g., running, walking). Ideally, metadata describing human motion for multimedia should also contain these high-level descriptors such that videos containing these motions can easily be searched by entering some simple keywords. It should not be too detailed and include the idiosyncrasies of an individual's movements (such as provided by motion capture data) but detailed enough to identify the moment of interest (e.g. when John in Section 1 jumps into a lake) so as to provide a precise temporal match of a motion. Motion descriptions such as LabanXML and MovementXML provides a direct translation from the underlying dance notation, and hence suffers from a searchability point of view due to their reliance on dance notation's musical concepts which generally does not have a corresponding counterpart in computer-based multimedia file formats. Although one can potentially describe human motion using a combination of MPEG-7 descriptors, the resulting descriptors would be complex and not conducive to a search environment where an efficient, simple, and scalable human motion description is required.

3. The human motion description format

3.1 Requirements

To answer the question of the missing verb in multimedia search as described in Section 1, a searchable Human Motion Description format would have to meet the following requirements:

1. *Compatibility with existing multimedia description formats*: The description would only need to serve as a bridge between highly temporal and non-temporal multimedia description formats for search purposes. Therefore, it needs to work together with other description formats, notably Dublin Core and MPEG-7.
2. *Limb based*: To describe human motion accurately, the movement of individual major limbs such as the arms and the legs has to be described. Further, the traditional human

motion annotation such as Labanotation is limb-based, proving the flexibility and accuracy of a limb-based approach. In terms of search, a search for a hand waving motion would result in a standing and waving motion and walking and waving motion, in effect expanding or restricting the detail level of the search according to the wishes of the user.

3. *Anatomical plane-based:* To describe a motion independent of the direction that the body is facing, the anatomical planes of the human body (Fig. 1) provides a fixed frame of reference relative to the body itself.
4. *Non-reconstruction*: Since the description would only describe general, temporal movements of the body, reconstruction for motion playback purposes (e.g. motion capture formats) is not needed. This requirement would prevent the description scheme to be too complicated and too detailed to search.
5. *Temporal*: It needs to be able to describe human motion temporally to provide a detailed description that can be used to search for a specific moment in time of a motion. E.g. it has to be able to search for a "*running then jumping*" motion, with a clear separation between the running and the jumping portions of the motion.

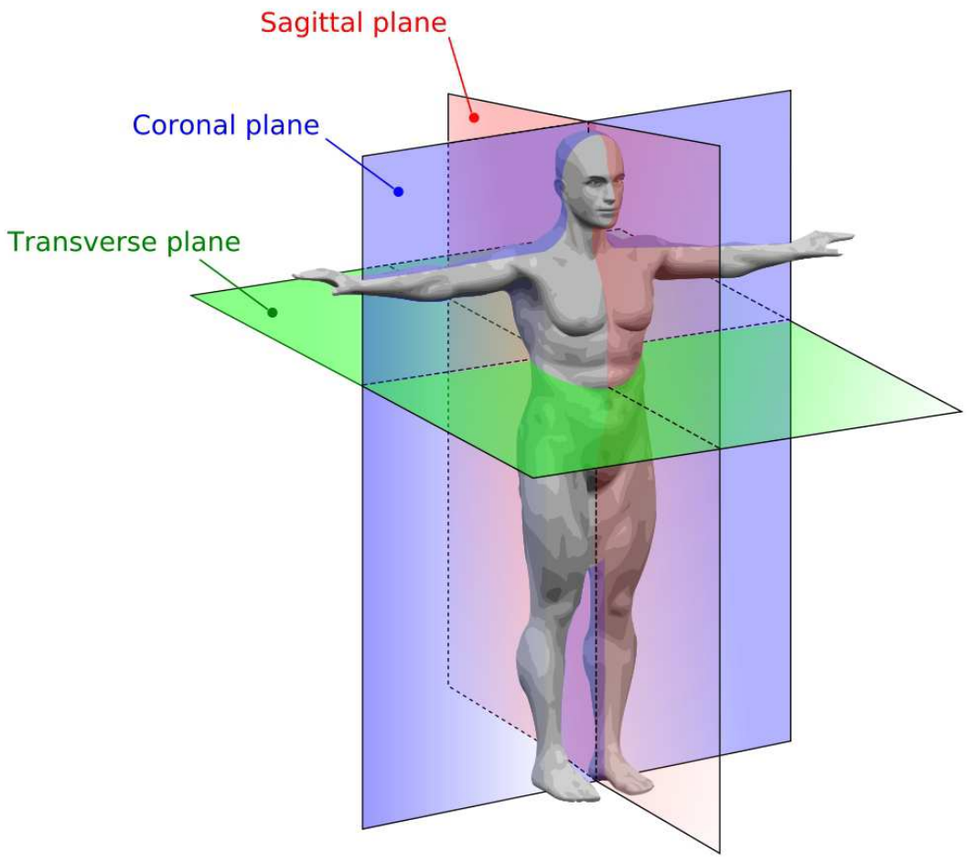


Fig. 1. Graphical representation of the anatomical planes that divides the body (Mrabet 2008) The Sagittal plane divides the left/right sections of the body, the Coronal plane divides the front/back sections of the body, and the transverse plane divides the top/bottom sections of the body. (Licensed under Creative Commons Attribution and ShareAlike license (CC-BY-SA)).

3.2 Design and XML schema visualization

Taking into account all the outlined requirements in Section 3.1, the Human Motion Description (HMD) format has been implemented as an XML Schema (see the visualization of the schema in Fig. 2). XML was chosen as it provides maximum compatibility between existing popular description formats - Dublin Core can be described using XML, and MPEG-7 is entirely XML-based. The core elements of the HMD schema is as follows (see Fig. 2):

- The element <motion> that encapsulates the motion description format, with an optional *fps* attribute that specifies the number of recorded frames per second in the annotated motion.
- The elements <sagittal>, <coronal>, and <transverse> that encapsulate descriptions in their respective axis:
  - The Sagittal plane divides the body between left and right.
  - The Coronal plane divides the body between front and back portions.
  - The Transverse plane divides the body between upper and lower portions.
- The elements<leg>, <arm> and <torso> describe the movements of the limbs, which contains optional attributes:
  - *side*: specifies either the left or the right side for the arms and the legs.
  - *dir*: describes the direction of the movement.
  - *frame*: describes the frame number in the media when the movement started.
  - *occur*: for repetitive motions, describes how many times the exact motion is repeated.

In HMD, the motion of a limb is described in three planes simultaneously (sagittal, coronal, and transverse). The full directional descriptors for each plane are shown in Table 1, and the relationships between elements in the XML Schema are shown in Fig. 2. In Fig. 2, each limb has three directional descriptors, with the restriction that only one direction is to be specified for a movement (since a limb cannot move in two directions in the same plane simultaneously). Using the plane-based description format, this enables HMD to describe 3D motion, even for describing a 2D video. For example in Fig. 2, the arms in the sagittal plane can move in forward, backward, and center directions. In the coronal plane in right, left, and center directions, and in the transverse plane in high, low, and level directions. Exceptions are the "*center*" direction that only applies to the sagittal and coronal plane, and the "*none*" direction that applies to all three planes.

Direction	**Meaning**		
	Sagittal	**Coronal**	**Transverse**
forward	Forward	-	-
backward	Backward	-	-
left	-	Left	-
right	-	Right	-
center	Toward neutral position	Toward neutral position	-
high	-	-	Up
low	-	-	Down
level	-	-	Middle position
none	Not moving/held in place	Not moving/held in place	Not moving/held in place

Table 1. Directions of limb movements according to the body planes.

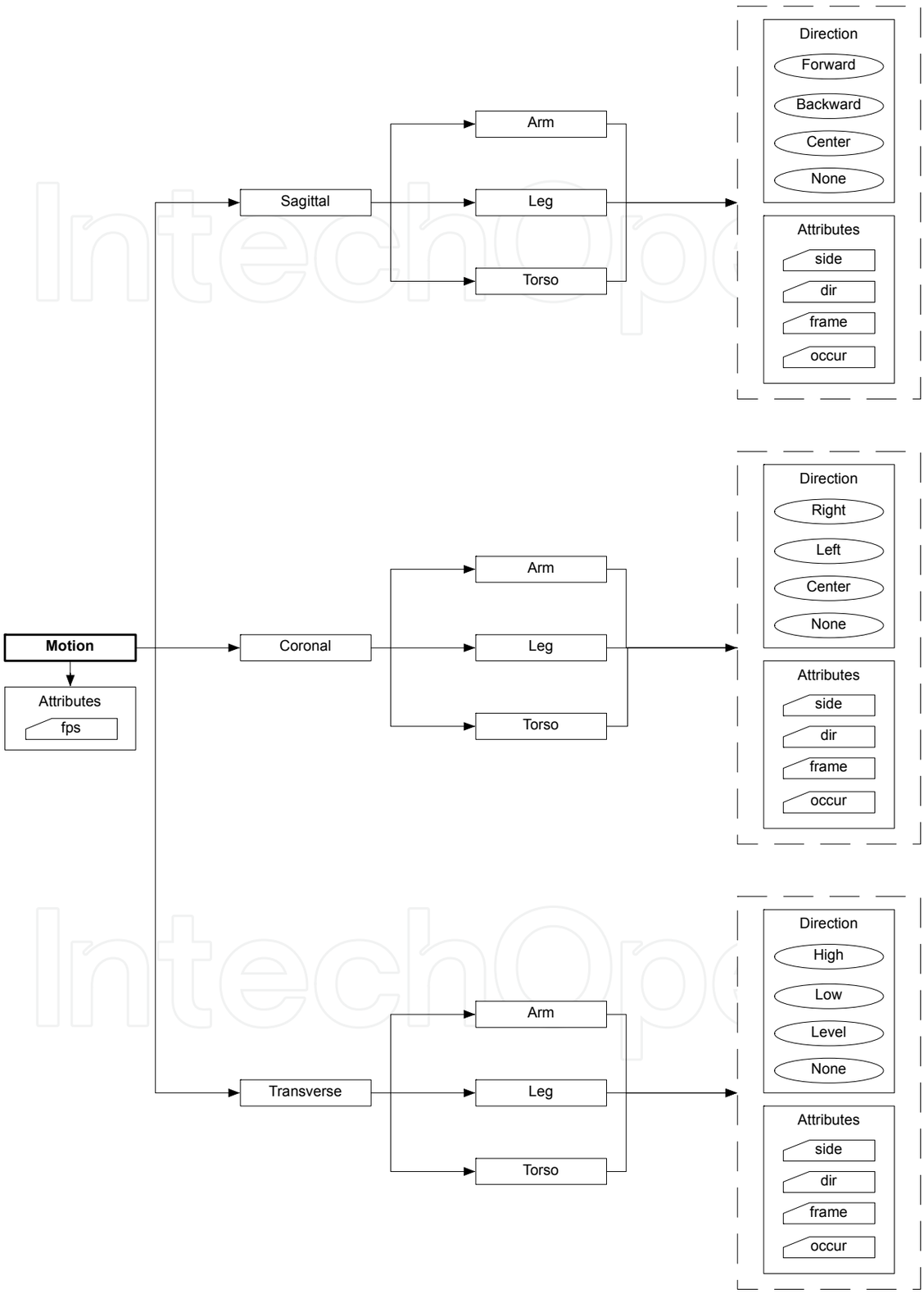


Fig. 2. Visualization of the Human Motion Description (HMD) schema element hierarchy.

3.3 Example: Using HMD to describe a motion

To illustrate the use of the HMD XML Schema as defined in Section 3.2, an example description using a walking motion is presented in Fig. 3 and Fig. 4. Fig. 3 illustrates a conceptual diagram of a HMD description of the leg movements in a walking forward motion. The sagittal, coronal, and transverse plane form individual description tracks, inside of which are the actual description of the limb movements. Specifying the frame number of the moment when the movement was made by the limb provides temporal information. The example in Fig. 3 depicts a four step walking motion starting with the right leg:

- The sagittal plane description track shows that the leg moves forward starting with the right leg.
- The coronal plane description track shows that the first movement of the right leg involves movement to the right, and the left leg movement involves movement to the left. In walking motion, the movement is slight, as the center of the mass of the body is moving according to the leg that is to receive the body weight in walking.
- The transverse plane description track shows that the legs move in a level direction.

In combination, the tracks provide a detailed description of the movements of the legs within the three body planes. The XML instantiation (based on the Schema) is shown in Fig. 4, where the *frame* attribute is present to denote the exact frame where each leg-forward motion is performed.

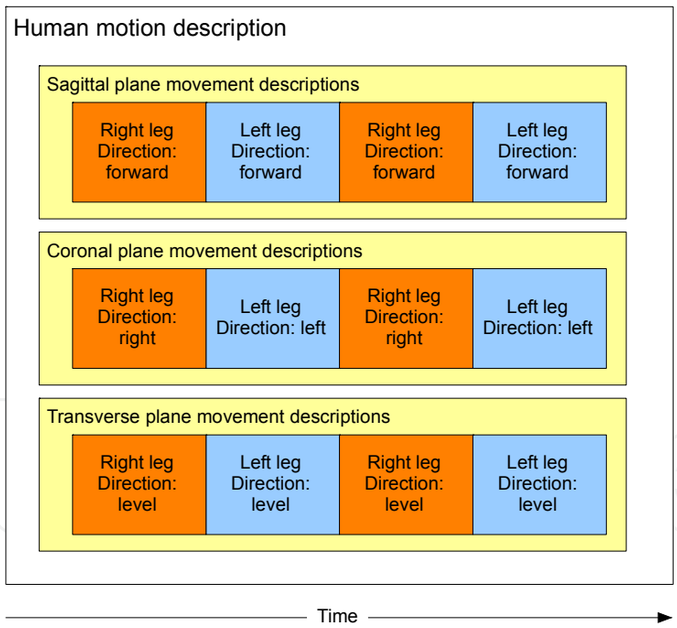


Fig. 3. An example diagram of the human motion description describing leg movement temporally in three separate planes.

Using the John Doe scenario in Section 1, a description of a running and jumping motion is shown in Fig. 5. In the running portion of the description, the leg forward motion appears to have exactly the same description as in the walking motion in Fig. 4. This is expected since both walking and running motion involves the same leg-moving-forward aspect. The difference is the timings involved. Note that in Fig. 4 there are 60 frames separations

between the motions, while in Fig. 5 the separation between motions is 40 frames. The shorter timing of the leg forward motion in Fig. 5 provides the differentiating factor between walking and running motions.

```
<hmd:motion fps="120">

  <sagittal>
    <leg side="right" dir="forward" frame="1" />
    <leg side="left" dir="forward" frame="60" />
    <leg side="right" dir="forward" frame="120" />
    <leg side="left" dir="forward" frame="180" />
  </sagittal>

  <transverse>
    <leg side="right" dir="level" frame="1" />
    <leg side="left" dir="level" frame="60" />
    <leg side="right" dir="level" frame="120" />
    <leg side="left" dir="level" frame="180" />
  </transverse>

</hmd:motion>
```

Fig. 4. Example XML instantiation from the HMD schema describing a walking forward motion.

```
<hmd:motion fps="120">

   <sagittal>

      <!-- run... -->
      <leg side="right" dir="forward" frame="1" />
      <leg side="left" dir="forward" frame="40" />
      <leg side="right" dir="forward" frame="80" />

      <!-- ...and jump (both feet off the ground) -->
      <leg side="left" dir="none" frame="120" />
      <leg side="right" dir="none" frame="120" />

   </sagittal>
   <transverse>

      <!-- run... -->
      <leg side="right" dir="level" frame="1" />
      <leg side="left" dir="level" frame="40" />
      <leg side="right" dir="low" frame="80" />

      <!-- ...and jump (both feet off the ground) -->
      <leg side="left" dir="none" frame="120" />
      <leg side="right" dir="none" frame="120" />
   </transverse>

</hmd:motion>
```

Fig. 5. Example XML describing John Doe's running and jumping motion using HMD.

Describing walking and running in a similar fashion provides an advantage: there is no need to provide separate descriptions for different motions that involves the same limb movement if the timings are all that differentiate the motions. In this way, walking, walking slowly, running, jogging, and sprinting can be described consistently and logically.

In the latter portion of John Doe's description in Fig. 5 marked as jumping, the jumping motion is described using the "*none*" attribute for the legs, which means that both feet are off the ground (simultaneously, in Fig. 5, as the motions for both legs occurred at the same frame number).

3.4 Example: Using HMD description fragments for a temporal query

While HMD is primarily a description format, fragments of HMD can be employed to perform temporal queries. An illustration of using a HMD fragment-to-HMD description query is shown in Fig. 6 by constructing parts of a walking motion, such as walking four steps in a forward direction starting with the right leg as viewed from the sagittal plane (from the side). The query portion of the illustration in Fig. 6 (with example XML fragment shown in Fig. 7) forms a subset of the walking motion description shown in Fig. 3 and Fig. 4. To perform a temporal search for a motion, the server would therefore search for a description that is the superset of the incoming query.

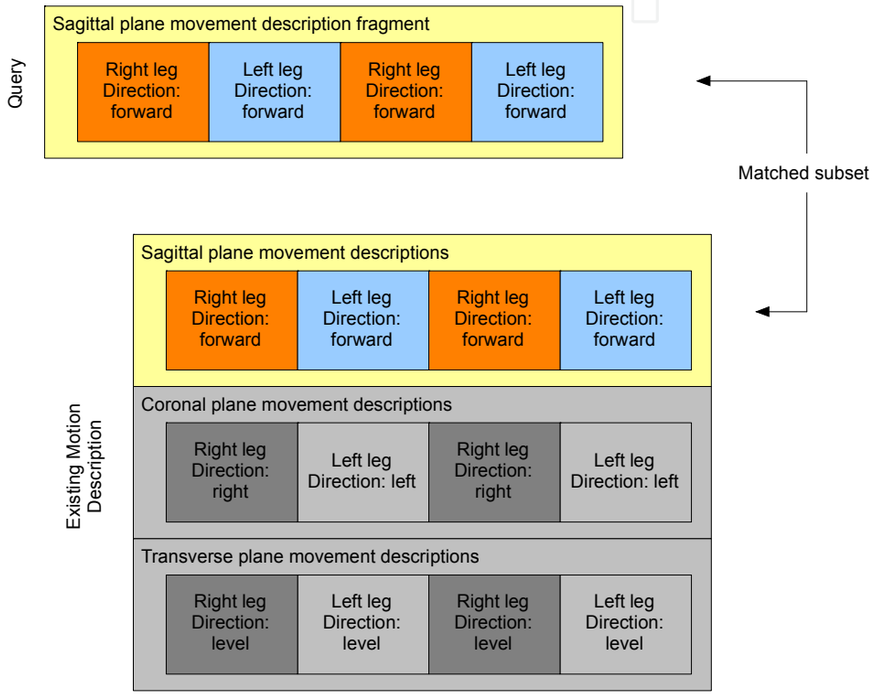


Fig. 6. An illustration of HMD temporal matching example by matching the incoming query to a subset of an existing description.

```
<hmd:motion fps="120">

  <sagittal>
    <leg side="right" dir="forward" frame="1" />
    <leg side="left" dir="forward" frame="60" />
    <leg side="right" dir="forward" frame="120" />
    <leg side="left" dir="forward" frame="180" />
  </sagittal>

</hmd:motion>
```

Fig. 7. Example XML of an HMD fragment forming a temporal query.

By matching a subset of an existing description, it is possible to perform a search using only a specific body plane, since the body planes divides the body in a constant manner no

matter the direction that the body is facing. For example, when searching a walking forward motion, the coronal plane movement is irrelevant. Similarly, when searching for a sidestepping motion, the movement of the sagittal plane is, in turn, irrelevant.

Another feature of HMD is the ability to match any number of repetitions of movement as detailed in Section 3.2. For example, to search for a walking forward motion, the query only contains the description of a leg forward movement with the "occur=*unbounded*" attribute. This signifies that the desired match is a leg forward motion that occurred for an undetermined number of repetitions. A block diagram of the "*unbounded*" attribute in use in a matching scenario is shown in Fig. 8. An XML instantiation example based on the HMD Schema is shown in Fig. 9. The query formed in Fig. 9 will match the example XML description shown in Fig. 4.

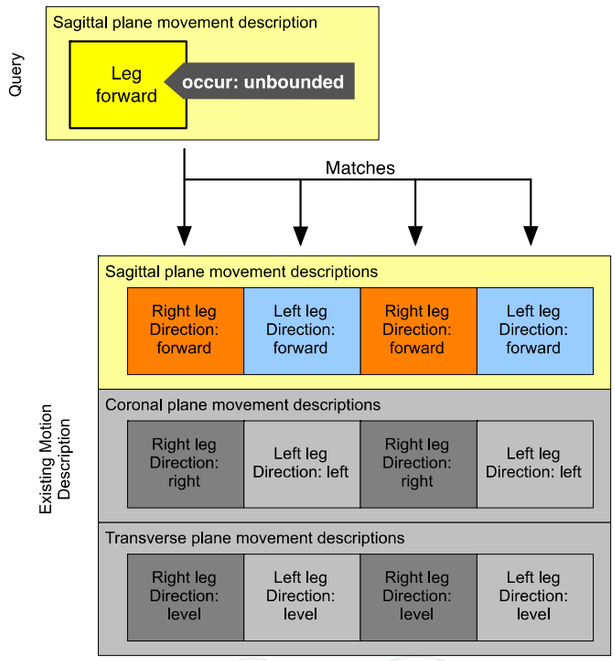


Fig. 8. An illustration of the "*unbounded*" occurrence of leg forward motion as a query term. The query term matches the same walking forward motion previously described.

```
<hmd:motion>

   <sagittal>
      <leg dir="forward" occur="unbounded" />
   </sagittal>

   <transverse>
      <leg dir="level" occur="unbounded" />
   </transverse>

</hmd:motion>
```

Fig. 9. Example XML fragment for using the "*unbounded*" attribute of HMD to form a query for multiple repetitive forward leg motions.

In the case of the John Doe example in Section 1, the running and jumping portion of the query is instantiated as shown in Fig. 10. In Fig. 10, the movements of the legs in the sagittal and transverse planes initially follow the description shown in Fig. 9. However, both legs are described to be using the "*none*" direction, which signifies that the feet are off the ground.

```
<hmd:motion>
   <sagittal>

      <!-- run... -->
      <leg dir="forward" occur="unbounded" />

      <!-- ...and jump (both feet off the ground) -->
      <leg side="left" dir="none" occur="1" />
      <leg side="right" dir="none" occur="1" />

   </sagittal>
   <transverse>

      <!-- run... -->
      <leg occur="unbounded" />

      <!-- ...and jump (both feet off the ground) -->
      <leg side="left" dir="none" occur="1" />
      <leg side="right" dir="none" occur="1" />
   </transverse>
</hmd:motion>
```

Fig. 10. XML instantiation of John Doe's running and jumping query from Section 1.

3.5 Example: Using HMD in conjunction with Dublin Core and MPEG-7 for search result filtering

Going back to the *Subject-Verb-Object* query concept described in Section 1, the search term "*John Doe running and then jumping into a lake*" can be thought of as a series of increasingly strict search filters.

Fig. 11 illustrates the flow of the filtering process involved in a detailed temporal search using human motion. The terms are separated according to their *Subject-Verb-Object* construct, with the subject described using Dublin Core (*John Doe*), the verb using HMD (*…running and jumping into…*), and the object using MPEG-7 (*…a lake*). Working together, the *Subject-Verb-Object* provides an increasingly strict filtering criteria for the media to be searched: Assuming that John Doe is searching for his own video that he uploaded to YouTube (100 of them in this scenario), he is searching for a video of himself running and jumping (there are 10 videos that matches the description), and only one of him actually jumping into a lake.

The example in Fig. 11 illustrates that HMD was intended to work together with existing multimedia description formats, the combination of which can provide a detailed temporal search using human motion.

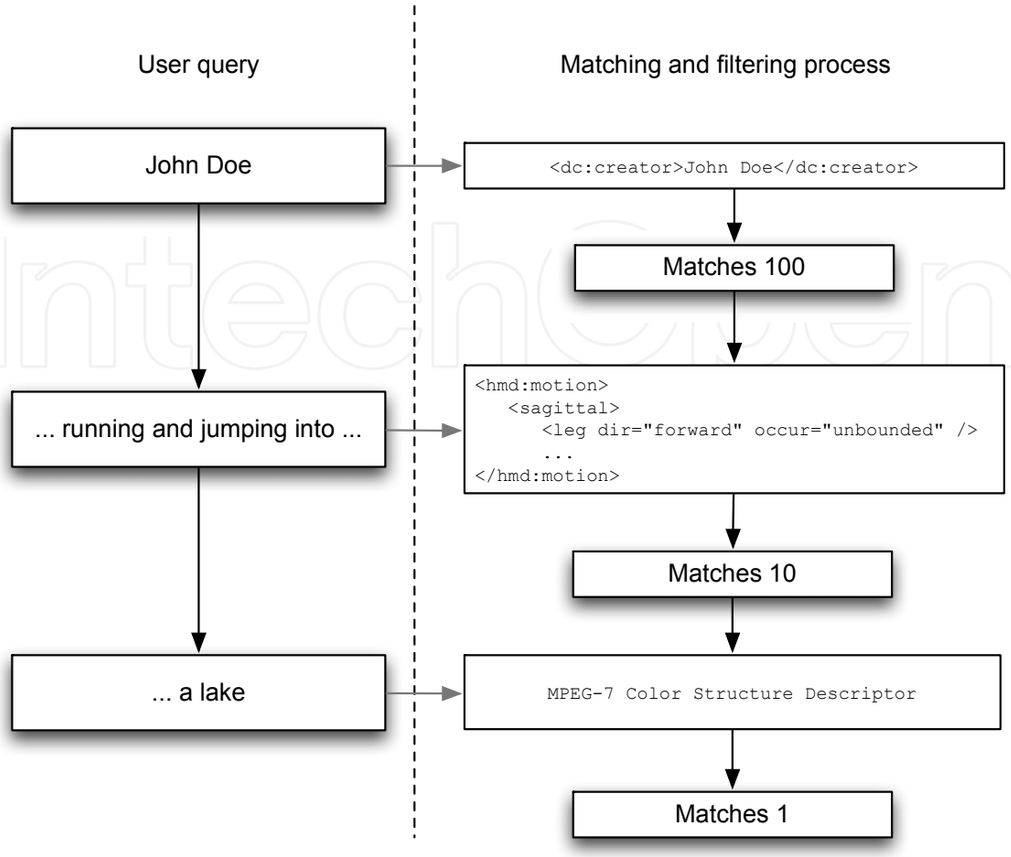


Fig. 11. An illustration of using HMD in conjunction with Dublin Core and MPEG-7 to perform a detailed search of a video using the query terms as an increasingly strict filtering process.

4. Conclusions and future work

This chapter has demonstrated a human motion description format designed for human motion enhanced multimedia search. Multimedia search using existing description formats can be thought of as increasingly strict search filtering based on a *Subject-Verb-Object* construct, where a non-temporal Dublin Core can describe the *Subject*, and temporal MPEG-7 descriptors can describe the *Object* portions of the construct. However, the important *Verb* portion is missing, which HMD intends to fill.

HMD was not designed to replace either Dublin Core or MPEG-7. Instead, it was designed to work in conjunction with both existing standards to provide a richer multimedia query environment, where a detailed query such as "*John Doe running and jumping into a lake*" can be performed with high temporal accuracy. HMD is based on XML Schema, which provide interoperability with any future description format that is based on XML. Key features of HMD include temporal matching for queries, limb-based description of motion using anatomical body planes that provides a constant frame of reference independent of the direction that the body is facing, and the ability to match arbitrarily repeated motions (such as leg movements in walking) using a single query term.

For future work, investigation into possible automatic extraction of HMD descriptions will be performed using a limb-based feature extraction as described in (Adistambha et al. 2011) as a basis. With automatic extraction, a user can upload a video to a video-sharing site such as YouTube, and using HMD, the uploaded video can be immediately indexed for detailed temporal searches involving human motions.

5. References

Adistambha, K. et al., 2011. Limb-based Feature Description of Human Motion. In International conference on signal processing and communication systems (to be presented). Honolulu, Hawaii.

Ann Hutchinson, 1970. *Labanotation; or, Kinetography Laban: the system of analyzing and recording movement*, Oxford University Press.

Barbič, J. et al., 2004. Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004*. Canadian Human-Computer Communications Society, pp. 185-194.

Gu, J. et al., 2010. Action and gait recognition from recovered 3-D human joints. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 40(4), pp.1021-1033.

Gutemberg Guerra-Filho & Yiannis Aloimonos, 2007. A Language for Human Action. *IEEE Computer Magazine*, 40(5), pp.60-69.

Hatol, J., 2006. *MovementXML: A representation of semantics of human movement based on Labanotation*. Simon Fraser University.

Laban, R. & Lange, R., 1975. *Laban's principles of dance and movement notation*, in 1956 and 1970 under title: Principles of dance and movement notation Includes indexes.

Li, C., Zheng, S.Q. & Prabhakaran, B., 2007. Segmentation and recognition of motion streams by similarity search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 3(3), p.16-es.

Loke, L., Larssen, A.T. & Robertson, T., 2005. Labanotation for design of movement-based interaction. In *Proceedings of the second Australasian conference on Interactive entertainment*. Sydney, Australia: Creativity \& Cognition Studios Press, pp. 113-120. Available at: http://portal.acm.org.ezproxy.uow.edu.au/citation.cfm?id=1109180.1109197 [Accessed December 10, 2009].

Martínez, J.M., 2004. MPEG-7 Overview. Available at: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm [Accessed October 9, 2011].

Mrabet, Y., 2008. *Planes of human anatomy*, Available at: http://commons.wikimedia.org/wiki/File:Human_anatomy_planes.svg.

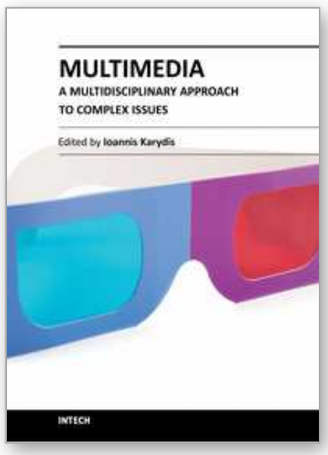
Mueller, M., Baak, A. & Seidel, H.-P., 2009. Efficient and Robust Annotation of Motion Capture Data. In ACM SIGGRAPH/Eurographics Symposium on Computer Animation.

Nakamura, M. & Hachimura, K., 2006. An XML representation of labanotation, LabanXML, and its implementation on the notation editor LabanEditor2. *Pregled Nacionalnog centra za digitalizaciju*, pp.47-51.

W3C, 2004. RDF Primer. Available at: http://www.w3.org/TR/rdf-primer/ [Accessed October 11, 2011].

Wei, X. & Chai, J., 2010. Videomocap: modeling physically realistic human motion from monocular video sequences. *ACM Transactions on Graphics (TOG)*, 29(4), pp.1-10.

Weibel, S. et al., 1998. Dublin core metadata for resource discovery.



Multimedia - A Multidisciplinary Approach to Complex Issues
Edited by Dr. Ioannis Karydis

ISBN 978-953-51-0216-8
Hard cover, 276 pages
Publisher InTech
Published online 07, March, 2012
Published in print edition March, 2012

The nowadays ubiquitous and effortless digital data capture and processing capabilities offered by the majority of devices, lead to an unprecedented penetration of multimedia content in our everyday life. To make the most of this phenomenon, the rapidly increasing volume and usage of digitised content requires constant re-evaluation and adaptation of multimedia methodologies, in order to meet the relentless change of requirements from both the user and system perspectives. Advances in Multimedia provides readers with an overview of the ever-growing field of multimedia by bringing together various research studies and surveys from different subfields that point out such important aspects. Some of the main topics that this book deals with include: multimedia management in peer-to-peer structures & wireless networks, security characteristics in multimedia, semantic gap bridging for multimedia content and novel multimedia applications.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kevin Adistambha, Stephen Davis, Christian Ritz, Ian S. Burnett and David Stirling (2012). Enhancing Multimedia Search Using Human Motion, Multimedia - A Multidisciplinary Approach to Complex Issues, Dr. Ioannis Karydis (Ed.), ISBN: 978-953-51-0216-8, InTech, Available from: http://www.intechopen.com/books/multimedia-a-multidisciplinary-approach-to-complex-issues/multimedia-search-using-human-motion

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen