Multiphase Action Representation for Online Classification of Motion Capture Data

Samer Salamah
Faculty of Computer Science, GDV Chemnitz University of Technology
Str. der Nationen 62
09111, Chemnitz, Germany
samer.salamah@s2008.tu-chemnitz.de

Guido Brunnett
Faculty of Computer Science, GDV Chemnitz University of Technology
Str. der Nationen 62
09111, Chemnitz, Germany
guido.brunnett@informatik.tu-chemnitz.de

ABSTRACT

In this paper we introduce a novel, simple, and efficient method for human action recognition based on a multiphase representation of human motion. An action is considered as a finite state machine where each state represents a primitive motion called motion phase, which is simply a sequence of poses with predefined common features. Spatial-temporal and postural features introduced in previous work are redefined by using only 3D joint positions for features extraction and are extended by involving the relative movement of the body end-effectors as new features. We developed a framework for modelling a given motion in the proposed motion model, whereupon we used this framework to create a model database of 25 different actions. Using this database we conducted a number of experiments on data obtained from several sources as well as on distorted data. The results showed that the presented method has high accuracy and efficiency. Additionally, it can work offline and online in real time, and can be easily adapted to work on 2D data.

Keywords

Human motion, motion capture, motion segmentation, motion classification, action recognition.

1. INTRODUCTION

Motion capture data is the basis for a realistic animation, but it is expensive to produce, therefore, the reusability of it is very important. However, this reusability demands that the motion capture data is good segmented and annotated. The segmentation into natural motion phases increases the reusability; however, the basis for this segmentation is the recognition of motion phases. Moreover, motion capture data is used in medicine for the analysis and examination of joint movement and rehabilitation procedures. These fields continuously produce large stores of data so that it is hard and tedious to retrieve a particular motion manually. Therefore, many methods have been developed for automatic search and retrieval in these stores. Of late, marker-less motion capture data has achieved significant improvement in accuracy, which enables it to be used in control and surveillance systems, as well as in the human–robot interaction field. This demands instantaneous and precise action recognition, which is what our presented method can do. Many works such as [Jin07a] and [Bar04a] successfully could reduce the high dimensionality of motion capture data without semantic lost. Additionally, some other works such as [Liu06a] and [Zha11a] could capture meaningful human motion with a reduced marker set. Inspired by such works, we develop a motion model that depends only on the movement of the actor's end-effectors and some basic postural features. We extend the features introduced in [Sal15a] so that any primitive motion can be described automatically in high-level terms. In general an action consists of several phases each of which is represented by a subset of these features and characteristics. Using the framework of phases and features a person with no experience with motion capture data is able to define or design movements at will and use them in any application area to retrieve and classify motions from motion repositories or to recognize ongoing motions online in real time.

The contribution of the proposed method is threefold: (1) specification of high-level features of human motion that enables (2) multiphase representation of human action and (3) utilizing this framework for efficient and high-accuracy classification of motion capture data. The presented approach is easy to implement, efficient, and works in real time both online and offline. Additionally, the database of recognizable motions can be extended easily in a very short time because there is no need for training data or training time. The rest of this paper is organized as follows: First, an overview of the

related works is given, and then some terms and notations used in our work are introduced. After that, the proposed features are described in Section 4 while the developed motion model is introduced in Section 5. In Section 6 the classification algorithm is presented, and then in Section 7 some conducted experiments are described and their results discussed. Finally, the work is concluded in Section 8.

2. RELATED WORK

Action recognition from motion capture data has received a lot of attention in the last decade. Nowadays there is a wide range of methods for classification of motion capture data. These methods can be divided into online and offline methods depending on whether the whole data should be processed before a classification result can be given or not. From another point of view, the classification methods can be divided into the following groups based on the nature of the features used to represent human motion as well as the field in which the used algorithms originated:

Description-Based

Methods of this category use annotated motion templates and high-level semantic features for action recognition. The work of J Baumann et al. [Bau14a] is an example of these approaches, where a motion capture database is annotated with actions of interest in an offline phase, and then used in the online phase to search for motion segments that are similar to annotated actions in the motion database. Leightley et al. [Lei14a] used Exponential Map EMP and k-means clustering to model human actions. For each action class they transform each pose of a representative sequence into EMP form then they used k-means clustering to extract a small number of exemplars that represent the action. Then they used Dynamic Time Warping and Template Matching to recognize actions from motion capture data streams.

Machine Learning

Machine learning techniques are widely used to classify 2D and 3D human motion. Cho and Chen [Cho13a] generated features for each motion frame based on the relative positions of joints, temporal differences, and normalized trajectories of motion. They then used them in training deep neural networks that they later used to classify motion capture data. Coppola et al. [Cop15a] extended the 3D Qualitative Trajectory Calculus (QTC3D) and used them to model human actions. Then they learned HMM to recognise human actions.

Statistics-Based

Statistical techniques such as Gaussian-Mixture-Models, Histograms and Space-Time Correlation are used here to model and recognize human motion. Y Jin and B Prabhakaran [Jin07a] quantized human motion data by extracting spatial–temporal features using SVD and then translated them into a one-dimensional sequential representation through a semantic Gaussian Mixture Models with Expectation-Maximization algorithm. These could reduce the dimensions of human motion data while maintaining semantically important features. M Zhang and A Sawchuk [Zha12a] introduced a framework for human motion modelling and recognition based on a bag of features. They modelled human activities through histograms of primitive symbols on physical features using k-means clustering and soft weighting. Unlike our proposed method, most of the above-mentioned methods are unable to separate two consecutive occurrences of one motion. In addition, the transitions between two motions are not recognized as transition but merged with the neighbour motions. Moreover, in some methods the learning process by classification is not simple, while our method is simple, easy to implement, efficient, and does not need any training phase.

3. PRELIMINARIES

We describe a pose of the human body as a set of annotated 3D points that correspond to the body joints. Thus, the human body pose is determined by the global 3D positions of these joints additional to the global orientation of the body. The proposed method needs a minimum set of joints *J*, namely, the ankles, knees, hips, chest, head, wrists, as well as a virtual joint at the pelvis called 'root'. In this work, we refer to ankles and wrists as feet and hands respectively. A pose at time *t* is described by $P^t = (q^t, p_1^t, p_2^t, \dots, p_n^t)$, where q^t is the global orientation of the body and p_j^t is the 3D global position of the joint *j*, where *n* is the number of used joints. The global body orientation at time *t* is given by three orthogonal vectors f^t, s^t, and h^t representing the normal vectors of the frontal, sagittal, and traversal main body planes respectively. We denote the single position coordinates of joint *j* at time *t* as x_j^t, y_j^t and z_j^t respectively where y_j^t is the vertical coordinate. We refer to the vector that goes from joint *a* to joint *b* at time *t* as $v_{a,b} = p_b^t - p_a^t$, and the motion direction of joint *j* at time *t* as $d_j^t = p_j^t - p_j^{t-1}$. Additionally, we define the motion magnitude of joint *j* at time *t* in the direction *v* as following $d_{j,v}^t = ||d_j^t|| \cos(\angle(d_j^t, v))$, where $v \in \{f^t, s^t, h^t\}$, and we refer to the algebraic sum $\sum_{t=s}^{t=e} d_{j,v}^t$ as the accumulated motion magnitude of joint j over the time interval $T = [s, e]$ in the direction *v*.

4. FEATURE DESCRIPTION

The main idea of the proposed method is based on a set of features that was inspired by the way in which people in general and kinesiologists in particular analyse and evaluate human motion. The method also

seeks to analyse the most important factors in deciding on the motion class. We extend the taxonomy tree of human motion introduced in [Sal15a] by adding motion directions of the end-effectors in the main body planes. The extended tree shown in Fig. 1 now consists of nine levels that reflect the importance of each group and the relations among features where the features in the first level have the highest importance. We call each complete path in this tree a 'pose state', which can be described as a complete set of the defined features. Each given pose is assigned a pose state by taking a previous pose into account. In the following, we introduce a detailed description of each of the features. In [Sal15a] the used features are calculated using both joint angles and 3D joint positions. However, we use here only 3D joint positions for calculating the introduced features.

Spatial–Temporal Features

In this section we introduce features that are generated by changing the joint positions over time, thereby denoting it as spatial–temporal features. They are introduced in the following in the order in which they are computed.

4.1.1 Motion Existence

First the existence of motion is checked. A pose is classified as dynamic if there is at least one joint that has moved a significant distance on at least one coordinate axis (1), otherwise it is classified as static.

$$\exists j \in J: \exists c \in \{x, y, z\}: \mid c_j^t - c_j^{t-1} \mid > \varepsilon \quad (1)$$

The threshold ε is a small real value representing the maximal noise value in the used data. Assuming there is a clip of *n* static poses that can be recorded during the system setup; the threshold ε is then the maximal displacement that a joint has achieved along any of the coordinate axes between two subsequent poses over the whole clip (2).

$$\varepsilon = \max_{t,j,c}(|c_j^t - c_j^{t-1}|) \text{ for all } t \in [2, n], \text{ all } j \in J \text{ and all } c \in \{x, y, z\}. \quad (2)$$

4.1.2 Motion Directions

Secondly, the motions of the end-effectors in the three main body planes are described. Based on the observation that almost all human actions are performed by displacing the body end-effectors, namely the hands, the feet, and the head/torso, we use the motion direction of these body parts as high-level features such as left foot moves forward up, or right arm moves left down fast. From a kinesiological perspective, the movements of body parts occur mainly in three anatomical planes, namely the frontal, sagittal, and traversal planes [Ham02a, Gre05a]. Based on this division of the body into three planes we define the directions of the joint movements relative to the body's axes as shown in Table 1.

Body Axis	frontal	vertical	sagittal
Positive Motion	forward	upward	left
Negative Motion	backward	downward	right

Table 1: Defined motion directions relative to main body's axes

4.1.3 Motion Space

Although the human body can move in many different ways, there are actually two major kinds of movements. These are locomotive, translator or linear, and non-locomotive, rotary, or angular [Ham02a, Gre05a]. If the whole body moves from one place to another, then the movement is locomotive; otherwise, it is considered as non-locomotive. A given pose is classified as locomotive if the root and both feet move, relative to the previous pose, in the same direction (3), or the root and at least one foot move in the same direction (4 and 5), while the other foot is fixed, and the accumulated magnitude of the root motion in the considered direction is greater than a certain threshold equal to the tibia length.

$$(\|d_{root}^t\| > \varepsilon) \wedge (\|d_{lfoot}^t\| > \varepsilon) \wedge (\|d_{rfoot}^t\| > \varepsilon) \wedge (d_{root}^t \cdot d_{lfoot}^t > 0) \wedge (d_{root}^t \cdot d_{rfoot}^t > 0) \quad (3)$$

$$(\|d_{root}^t\| > \varepsilon) \wedge (\|d_{lfoot}^t\| > \varepsilon) \wedge (\|d_{rfoot}^t\| \leq \varepsilon) \wedge (d_{root}^t \cdot d_{lfoot}^t > 0) \quad (4)$$

$$(\|d_{root}^t\| > \varepsilon) \wedge (\|d_{lfoot}^t\| \leq \varepsilon) \wedge (\|d_{rfoot}^t\| > \varepsilon) \wedge (d_{root}^t \cdot d_{rfoot}^t > 0) \quad (5)$$

where ε is the noise threshold defined in (2).

Postural Features

An important factor for classifying human motion is the change in the main body posture. We utilize this observation and use the following major and corresponding minor postures as features for the recognition of human actions.

4.1.4 Standing

In general, 'standing' is a major posture where the body maintains an upright position supported by the feet. The presented approach restricts the upright constraint to the lower body. Therefore, a pose is considered as 'standing' if at least one leg is extended and has a certain maximum inclination (6).

$$\left(\|v_{lfoot,lhip}\| > \eta\right) \wedge \left(\angle\left(v_{lfoot,lhip}, OY\right) \leq \alpha\right)) \vee \left(\|v_{rfoot,rhip}\| > \eta\right) \wedge \left(\angle\left(v_{rfoot,rhip}, OY\right) \leq \alpha\right)) \quad (6)$$

We consider a leg as extended if the distance between the foot and hip is greater than η, which is equal to one and a half of the femur length.

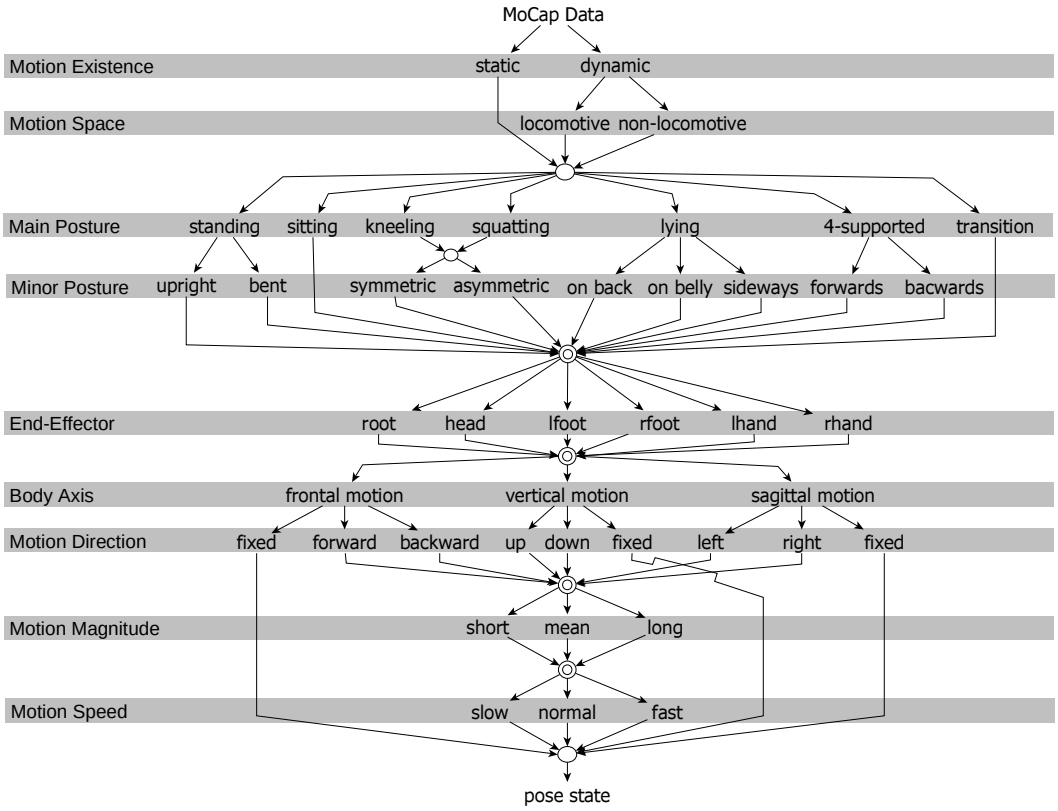

Figure 1: Taxonomy tree of human motion. Double circles allow the path to return to the first previous double circle, whereby it is not allowed to take the same path segment again.

The maximum inclination used by our experiments is $\alpha = 45°$. Standing can also have one of the following three minor postures:

1. If the torso stays upright, i.e. it has an inclination smaller than threshold β: $\angle(v_{root,cheast}, OY) \leq \beta$ (7), then the pose is considered as 'standing upright'. We used $\beta = 30°$.
2. Otherwise it is considered as 'standing bent': $\angle(v_{root,cheast}, OY) > \beta$. (8)
3. If the body is not supported only by the feet, then the pose is considered as 'standing leaned'. Suppose *S* is the set of support body parts, then 'standing leaned' is recognized when *S* contains at least one part except the feet $S \setminus \{p^t_{lfoot}, p^t_{rfoot}\} \neq \emptyset$. This minor posture, however, is in our case not recognizable, because motion capture data does not contain any information about the environment.

4.1.5 Sitting

The 'sitting' posture is a major posture in which the body is supported mainly by the buttocks rather than the feet, that implies that the projection of the gravity centre of the body lies outside the support base of the body formed through the feet. Additionally, the torso is not horizontal. Based on the height of the hip joint, it is decided whether the pose is sitting on an object or on the floor as minor postures. No constraints are put on the legs because there are many variants of the sitting posture according to the position of the legs. Legs can be vertical, crossed, or on each other.

4.1.6 Kneeling

'Kneeling' is also a major body posture in which at least one knee touches the ground and the root height is greater than half of the femur length, which is denoted as δ in (9). If only one knee fulfils these criteria, then kneeling is called asymmetric; otherwise, it is symmetric kneeling.

$$((y^t_{lknee} \approx y_0) \vee (y^t_{rknee} \approx y_0)) \wedge ((y^t_{root} - y_0) > \delta) \quad (9)$$

Given that the ground height can be greater than zero (stairs case), we denoted the ground height as y_0.

4.1.7 Squatting

'Squatting' is a major human body posture in which at least one foot touches the ground but not the knee, and the vertical distance between the corresponding hip and foot is smaller than half of the femur length (10). Additionally, the torso must not be horizontal.

$$((y^t_{lfoot} \approx y_0) \wedge (y^t_{lhip} < \delta) \wedge (y^t_{lknee} > y_0)) \vee ((y^t_{rfoot} \approx y_0) \wedge (y^t_{rhip} < \delta) \wedge (y^t_{rknee} > y_0)) \quad (10)$$

Squatting is symmetric when both the knees are bent; it is asymmetric when only one knee is bent.

4.1.8 Lying

'Lying' is a major posture in which the body is in a horizontal or resting position supported along its length. In the proposed approach, this definition is restricted to the torso, i.e. the torso should have an inclination greater than a threshold γ: $\angle(v_{root,cheast}, OY) > \gamma$ (11), which we set at 70° in the conducted experiments. If at least one hip lies on the floor, then the pose is classified as lying on the ground (12), otherwise on an object.

$$(y^t_{lhip} \approx y_0) \vee (y^t_{rhip} \approx y_0) \quad (12)$$

If the two hip joints have approximately the same height (13) and the normal of the frontal plane points down (14), then the pose is lying on the belly. If the mentioned normal points up (15) and the two hip joints have approximately the same height, then the pose is called lying on the back.

$$|y^t_{lhip} - y^t_{rhip}| < \|v_{rhip,lhip}\|/2 \quad (13)$$

$$\angle(f^t, OY) \approx 180° \quad (14)$$

$$\angle(f^t, OY) \approx 0° \quad (15)$$

If the difference between the heights of both the hips is greater than half of the distance between the two hip joints (16), then the pose is lying sideways.

$$|y^t_{lhip} - y^t_{rhip}| \geq \|v_{rhip,lhip}\|/2 \quad (16)$$

4.1.9 Four-Supported

In this rare major posture, the hands and the feet contact the ground but not the root $(y^t_{lfoot} \approx y_0) \wedge (y^t_{rfoot} \approx y_0) \wedge (y^t_{lhand} \approx y_0) \wedge (y^t_{rhand} \approx y_0) \wedge (y^t_{root} > y_0)$. If the belly faces the ground (14), then the position is called 'forward four-supported', or else the back faces the ground (15) and the position is called 'backward four-supported'. Another variant of this posture is when at least one upper limb and one lower limb contact the ground at the same time (17). This variant allows more movements to be performed than the first variant.

$$\left((y^t_{lfoot} \approx y_0) \vee (y^t_{rfoot} \approx y_0)\right) \wedge$$
$$\left((y^t_{lhand} \approx y_0) \vee (y^t_{rhand} \approx y_0)\right) \quad (17)$$

4.1.10 Transition

The transitions between the above-mentioned main postures of the human body are considered here. If the pose cannot be classified as one of the above-mentioned major or minor human body postures, then it is considered a transition posture. The previous and next major postures determine the name of the transition, i.e. the classification of a transitional posture is dependent on the two surrounding main postures. For example, the pose that corresponds to the transitional phase between 'sitting' and 'standing' will be classified as 'standing up'.

5. MULTIPHASE REPRESENTATION OF MOTION

Any human activity can be generally divided into a sequence of simple motions called 'phases'. This division makes the action classification easier and more robust. In the kinesiological analysis of human motion, one tries to divide the considered activity into three phases: preparatory phase, power phase, and follow-through phase [Ham09a], or preparation phase, action phase, and recovery phase [Bar07a]. Here each phase can be further divided into sub-phases so that each sub-phase consists only of some basic joint movements in the directions introduced in Section 4. We use, however, a certain definition of the motion phase and do not distinguish between power phase and other phases. We define the motion phase as a sequence of poses with a common set of features defined above in Section 4. Table 2 summarizes the feature set and the range of values of each feature, where the feature value 'undefined' denotes that this feature is not important in the considered phase, i.e. it can be ignored.

Feature	Values	
Motion Existen	static, dynamic, undefined	
Motion Space	locomotive, non-locomotive, undefined	
Major Posture	standing, sitting, kneeling, squatting, lying, four-supported, transition, undefined	
Minor Posture	standing	upright, bent, leaned, undefined
	sitting	on object, on floor, undefined
	kneeling	symmetric, asymmetric, undefined
	squatting	symmetric, asymmetric, undefined
	lying	{on belly, on back, sideway, undefined } ×{on object, on floor, undefined }
	four-supported	backwards, forwards, undefined
	undefined	
Frontal Motion	{forwards, backwards, fixed, undefined} ∪ $M \times S$	
Vertical Motion	{up, down, fixed, undefined} ∪ $M \times S$	
Sagittal Motion	{left, right, fixed, undefined} ∪ $M \times S$	

Table 2: Summary of introduced features and their possible values, where × stands for the Cartesian product operation, M = {short, mean, long, undefined} and S = {slow, normal, fast, undefined}

This definition of the wide range of high-level features allows the description of the most common human activities in a high language, enabling a comfortable retrieval system. Often, an action that consists of several phases can only be performed starting from a certain phase. In these cases the motion description involves the order of phases. On the other side there are some actions that can be started in more the one phase, such as the kicking action, which consists of three phases and can be started in the first or second phase, where in the first phase the used leg moves backwards to give the strike more power, then it moves forward long fast in the second phase and then moves backwards down to the rest position in the last phase. Here the first phase is optional because kicking can be performed without this phase. Table 3 shows the detailed definition of kicking using the right leg without the optional phase.

Feature	Phase 1	Phase 2
Motion Existence	dynamic	dynamic
Motion Space	non-locomotive	non-locomotive
Main Posture	standing	standing
Minor Posture	undefined	undefined
root Frontal-Vertical-Sagittal Motion	fixed-fixed-fixed	fixed-fixed-fixed
torso Frontal-Vertical-Sagittal Motion	undefined-undefined-undefined	undefined-undefined-undefined
lfoot Frontal-Vertical-Sagittal Motion	fixed-fixed-fixed	fixed-fixed-fixed
rfoot Frontal-Vertical-Sagittal Motion	forward long fast-up mean fast-fixed	backward long fast-down mean fast-fixed
lhand Frontal-Vertical-Sagittal Motion	undefined-undefined-undefined	undefined-undefined-undefined
rhand Frontal-Vertical-Sagittal Motion	undefined-undefined-undefined	undefined-undefined-undefined

Table 3: modeling the motion class "KickR" using the proposed motion model.

Another relative complex example is the jumping action. Jumping can be divided into four phases. In the first phase the feet stay fixed while the root moves down. In the second phase the whole body moves up and forwards, while it goes on forward in the third phase but down. In the last phase the feet are fixed while the root moves up and forwards.

6. ACTION RECOGNITION

Actions to be recognized should be manually modelled and saved in a model database using the developed framework. For each action in the action model database, a finite state machine FSM is created automatically (Fig. 2).

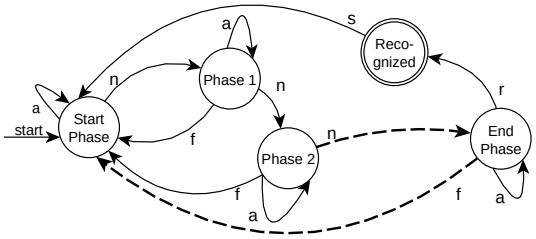

Figure 2: Finite state machine representing the recognition process of a defined action, where 'n' means that the next phase is matched and the current phase can be ended; 'a' means the current phase is matched; 'f' implies failed to match either the current phase or the next one; 'r' means the end phase ended successfully and the motion is recognized; 's' means return to the start phase and start again.

Suppose that an action model consists of n phases $S_1, S_2, \dots, S_n$, where $S_1$is the start phase and S_n is the end phase, then the corresponding FSM is defined as following: $A = (\Sigma, S, s_0, \delta, F)$, where Σ is the input alphabet and consists of all possible pose states; $S = \{S_1', S_1', \dots, S_n'\}$ is the states set and it consists of the action phases whereby each phase is extended to have the following attributes: (1) start time τ, (2) end time σ and (3) an activation flag. s_0 is the initial phase. δ is the transition function and it will be defined later in Fig. 3. F is the set of final states and it consists here of the extended end phase. The input data in each frame consists of the global positions of the used joints as well as the global body orientation. The motion features are computed using this information and then the FSM for each action is updated using the computed current pose state as shown in Fig. 2 and Fig. 3. At the beginning all created FSMs are considered to be in their initial phase. When a new pose is available, the pose state is computed and given to each FSM to update its status as following: if the pose state is compatible with the current FSM phase i.e. the phase is matched, then the phase is retained and the related action is considered active. Otherwise, if the current phase is not matched and it was active in the previous frame, then the phase is considered to be achieved and can be ended if the accumulated motion magnitude and motion speed of each required phase feature are within the desired range and, in this case, the FSM is aggregated to the next phase. Otherwise, the action is cancelled and the FSM is returned to its start phase. If it is assumed that S_t is the pose state of the pose *t*, i.e. S_t is a complete set of the defined features or a complete path in the taxonomy tree, and S_φ is the feature set of the current phase S_φ' of the FSM for the action $\mathcal{M}$, then the global recognition algorithm of the action $\mathcal{M}$ at the time *t* can be stated as follows:

```
1  if S_ω ⊆ S_t then
2      if the current action phase S'_φ is active then
3          set end time of S'_ω  σ = t.
4      else
5          set start time of S'_ω  τ = t.
6          raise the activation flag of S'_φ, i.e. make S'_φ active.
7  else if S'_ω is active and can be ended then
8      if the S'_φ is the end phase then
9          action M is recognized.
10         return to the first phase and reset the activation flag of all phases.
11     else move to the next phase.
12 else return to the first phase and reset the activation flag of all phases.
```

Figure 3: Transition function of the action FSM.

The proposed approach can provide information about the ongoing activity before it is completed, which is an important issue for some application areas such as human–robot interaction, because it enables the robot to response quickly and at the right time.

7. EXPERIMENTAL RESULTS

We developed a framework for action design and action classification from different motion capture databases, namely CMU [Cmu14a], HDM05 [Mue07a], and locally captured data (at our institute). The used data contains distorted walking data. Using our framework we modelled 25 actions manually as explained in section 5. The motion clips were first manually segmented and annotated by two different persons, and then processed by our system. Table 4 shows the actions used in our experiments and the measured evaluation values, where the global precision is about 96.2% and the global recall is more than 98.1%. To begin with, we measured the precision of action recognition as follows: *precision = count of correctly recognized action / count of all recognized actions*. Another evaluation value is the *recall*, which is *the percentage of the count of correctly recognized actions compared to the count of ground truth actions*. An action is considered correctly recognized if the temporal overlap between it and a manually segmented action of the same type is bigger than half the length of the manual action. We measured also the segmentation error as follows: the segmentation error is zero if the difference between the automatic detected cut and the manually created cut smaller than ten, otherwise the segmentation error is equal to this difference minus ten, where a manual created cut is the mean of all manual created cuts (in our case two) of the considered action. The proposed method is able to recognize some particular information about the action such as the marching foot while walking and running, the used hand while punching, or the leg while kicking. All occurrences of most of the defined actions are recognized successfully. An exception is the activity of walking. This is because sometimes the first and last strides of running are recognized as walking. The method failed to match the second phase in the running motion if the feet are not far enough from the ground. This is, however, a minor drawback, because walking and running are similar motions especially in terms of the first and last running strides.

Action Class	Prec-ision	Re-call	Segmentat-ion Error
WalkL	0.94	0.99	2
WalkR	0.93	0.99	1
RunL	0.98	0.94	0
RunR	1	0.96	0
BoxL	1	0.96	11
BoxR	0.96	1	19
KickR	1	1	10
KneeKickR	1	1	27
SideKickR	1	1	23
Jump	1	1	11
JumpJacks	1	1	7
StandUp	1	1	55
SitDown	1	1	14
Hop2Legs	1	1	71
HopR	1	1	31
HopL	1	1	20
SwingArmsSagittal	1	1	11
SwingArmsTravers	1	1	26
SwingArmsCircular	1	0.94	14
ChoppingL	1	1	4
ChoppingR	1	1	19
Fight	1	1	28
DrinkR	1	1	18
Throw	1	1	57
Squat	1	1	32

Table 4: Results of the experiments, where 'L' stands for left and 'R' for right and it refers to the active limb during the action.

The classification speed is linear with the number of actions to be recognized. The mean recognition speed for a model database of 25 actions amounted ~1200 fps on a computer running Windows 8 with AMD A4-4300M APU processor, 2.50GHz and 4.00GB RAM. If the database were hypothetically extended to contain 250 actions, then the speed would sink to ~120 fps. This means that our method can scale to large model databases and can still perform well in real time.

Compared to some other works which were evaluated using data from the same data sources which we used ,namely the HDM05 and CMU, the proposed method produces better results as shown Table 5. However this comparison might be unfair because the used datasets might be slightly different and the classes and numbers of considered actions are also different.

Action Class	[Cho13a]	[Lei14a]	[Zha12a]	Proposed
All	0.95	0.9492	0.927	0.962
Walk	-	~0.975	0.923	0.935
Run	-	~0.975	0.989	0.99
Hop	-	~0.95	1	1
Box	-	~0.86	-	0.98
Squat	-	~0.94	-	1

Table 5: Precision of some other works where „-“ stands for unknown accuracies and „~“ stands for those read from a diagram picture.

8. CONCLUSION AND FUTURE WORK

In this paper a set of high-level semantic features are introduced and employed in a multiphase motion representation that enables an efficient recognition and retrieval of motion capture data with high accuracy. The introduced features as well as the multiphase representation of motion are inspired by kinesiology, and hence the proposed method mimics the human mind by motion perceiving and analysing what enables it to perform very well. It can also work online and offline in real time. The recognizable motion database can be extended easily and in a short time, because our method does not require any training time. The experiments made on large databases from different sources, as well as on distorted data, proved that the proposed method scales well to other data sources. As future work we plan to extend this method so that it can also classify single poses, static clips, and static gestures.

9. ACKNOWLEDGEMENTS

Some of the datasets used in this work were obtained from mocap.cs.cmu.edu while some other sets were obtained from HDM05.

10. REFERENCES

[Bar04a] Barbic, J., Safonova, A., Pan, J. Y., Faloutsos, C., Hodgins, J. K. and Pollard, N. S.. Segmenting motion capture data into distinct behaviours. Graphics Interface, 185-194 , 2004.

[Bar07a] Bartlett, R.. Introduction to Spotrs Biomechanics: Analysing Human Movement Patterns 2nd Edition. ISBN 0-203-46202-5, Routledge, UK, USA and Canada, 2007.

[Bau14a] Baumann, J., Wessel, R., Krüger, B. and Weber, A.. Action Graph: A Versatile Data Structure for Action Recognition. International Conference on Computer Graphics Theory and Applications, 2014.

[Cho13a] Cho, K. and Chen, X.. Classifying and Visualizing Motion Capture Sequences using Deep Neural Networks. arXiv preprint arXiv:1306.3874, 2013.

[Cmu14a] CMU Graphics Lab Motion Capture Database, http://mocap.cs.cmu.edu/search.php?subjectnumber=86. Date of Access March, 16, 2016.

[Cop15a] Coppola, C., Martinez Mozos, O. and Bellotto, N.. Applying a 3D qualitative trajectory calculus to human action recognition using depth cameras. IEEE/RSJ IROS Workshop on Assistance and Service Robotics in a Human Environment, 2015.

[Gre05a] Greene, D. P. and Roberts, S. L.. Kinesiology: Movement in the Context of Activity 2nd Edition. ISBN 0-323-02822-5, Elsevier Inc., USA, 2005.

[Ham02a] Hamilton, N. and Luttgens, K.. Kinesiology: Scientific Basis of Human Motion 10th Edition. ISBN 0-07-112243-5. McGraw-Hill, USA, 2002.

[Ham09a] Hamill, J. and Knutzen, K. M.. Biomechanical Basis of Human Movement 3d Edition. ISBN-13: 978-0781791281 ISBN-10: 0781791286, Lippincott Williams & Wilkins, USA, 2009.

[Jin07a] Jin, Y. and Prabhakaran, B.. Semantic Quantization of 3D Human Motion Capture Data Through Spatial-Temporal Feature Extraction. MMM 2008, LNCS 4903, pp. 318–328, 2007.

[Lei14a] Leightley, D., Li, B., McPhee J., Hoon Yap, M., Darby, J.. Exemplar-Based Human Action Recognition with Template Matching from a Stream of Motion Capture. 11th International Conference, ICIAR 2014, Vilamoura, Portugal, 2014.

[Liu06a] Liu, G., Zhang, J., Wang, W. and McMillan, L.. Human Motion Estimation from a Reduced Marker Set. Proceedings of the 2006 symposium on Interactive 3D graphics and games, ACM, USA, 2006.

[Mue07a] Müller, M., Röder, T., Clausen M., Eberhardt, B., Krüger, B. and Weber, A.. Documentation Mocap Database HDM05 (Part-Scene #1). Germany, 2007.

[Sal15a] Salamah, S., Zhang, L., Brunnett, G.. Hierarchical Method for Segmentation by Classification of Motion Capture Data. Virtual Realities pages 169-186, ISBN 978-3-319-17043-5, 2015.

[Zha11a] Zhang, L., Brunnett, G. and Rusdorf, S.. Real-time Human Motion Capture with Simple Marker Sets and Monocular Video. Journal of Virtual Reality and Broadcasting, Volume 8, no. 1, 2011.

[Zha12a] Zhang, M. and Sawchuk, A. A.. Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach. Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, USA, 2012.