

Human Behavior Recognition Method based on Image Sequences

Yanhua Chen

*Modern education technology center, Neijiang Normal University,
NeiJiang 641100, china
cyhzb@126.com*

Abstract

Most researches on human behavior recognition are mainly based on the features of whole body motion. This paper proposed a hierarchical discriminative approach for recognizing human behavior based on limbs motion. The approach consists of feature extraction with mutual motion pattern analysis and discriminative behavior modeling in the hierarchical manifold space. A cascade CRF is introduced to estimate the motion patterns in the corresponding manifold subspace, and the trained SVM classifier is used to predict the behavior label for the current observation. The results on motion capture data prove the significance motion analysis of body parts, and the results on synthetic image sequences are also presented to demonstrate the robustness of the proposed algorithm.

Keywords: *human behavior analysis, image sequence, support vector machine*

1. Introduction

Of all studies on single behavior recognition, most researches focused on daily behaviors of people, such as walking, running, jumping, sitting, lying, shaking [1-3]. Those movements are apparently different in either 2D or 3D postures from the aspect of a whole human body or parts. Man's many behaviors are distinguishable maybe for only one or several human body parts. In such cases, it's very difficult to recognize variations of each behavior. What we'll discuss here is exactly recognition of such behaviors, *e.g.*, walking, goose-steps, tapping the ball, walking with still hands, hyperbolic walking. The work has lots of promising applications in scenes where robust human posture information is effectively acquired, *e.g.*, sport behavior analysis, digital amusement and virtual reality [4-5].

The single behavior recognition method proposed in the paper doesn't consider the extraction of low-level features; instead, it uses directly motion capture data and behavioral data based on synthetic human body pictures [6-7]. Although the information about 3D joint positions is known, it's still greatly hard to perform behavioral modeling and recognition in high-dimension body posture space. That is because in such a space, the efficiency of time for behavior reasoning and modeling is very low; besides, some key attributes of human behaviors are hardly observed. However, due to restriction of physical biomechanics and other movements, human motions can actually be regarded as movements in some low-dimension manifold space [8-9]. Hence, we employ hierarchical Gaussian process latent variable model to create the low-dimension latent variable space of human motions. We represent the behavior of whole human body by fetching the motional mode of each separate human body part from the related separate latent variable space. The feature expression method can help find the internal essence of human movements [10].

Each frame of human behavior sequence relates to one posture of human body. The pose sequence is expressed as different behaviors as per temporal expansion. HMM and the like was mostly-used and popular time sequence data modeling approach [11]. But for

generative models like HMM would require conditions of observed features should be mutually independent, hard to characterize the correlation of human behaviors in long time frame; while discriminative models require no assumption that conditions of observational process are independent, making more naturally the model of temporal-spatial neighborhood relationship of human behaviors. The paper develops one cascaded conditional random field (CRF) model to build the relationship between observed characteristics and partial body motion patterns. Then, with discriminative classifier SVM, we can identify the final behavioral tag. Experiments prove that for subtle single behavior like walking, walking with still hands, hyperbolic walking, the proposed algorithm shows satisfactory recognition performance. Meanwhile, we utilize 3D rendering software to establish single behavior database based on composite images. Experiments with the database validate that in spite of certain errors in 3D human posture based on the estimation of composite images, the technique shows strong robustness.

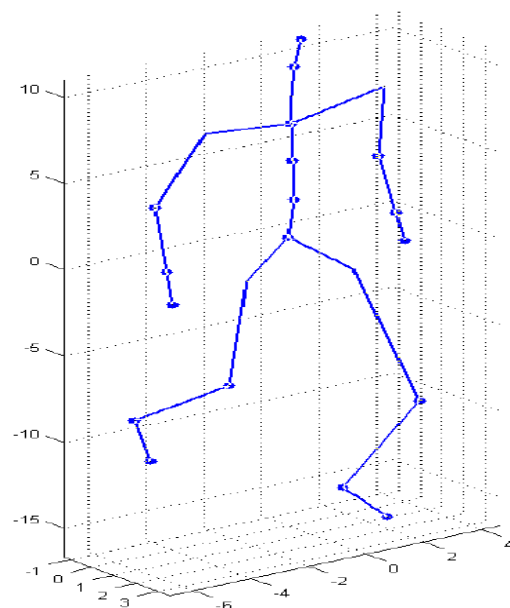
2. Motion Pattern

2.1. Hierarchical Latent Variable Space of Human Behaviors

As per common senses in biomechanics, we perform hierarchical decomposition of human body state space. With Hierarchical Gaussian Process Latent Variable Model, we build tree-like hierarchical manifold space of human movements. It is shown in figure1.

The tree-like manifold space requires only leaf nodes related to observed values of state space, with high-level latent variables to make model of the internal association of each low-level sub-manifold space, which defines well the condition independence of different movements of human body parts. To manifest the given joint probability distribution in Figure 1(b), we choose one typical tree-like branch as instance (*i.e.*, nodes in dotted box in Figure 1(b)). Suppose Y_{rl} and Y_{ll} corresponding to observation data of left and right leg movements; X_{rl} and X_{ll} are latent variables of left and right leg motions in low-dimension space; X_{lower} corresponding to low-dimension latent variable of father node movements. Then the given joint probability distribution of hierarchical latent variable space with three nodes can be defined as:

$$p(Y_{rl}, Y_{ll}) = \int p(Y_{rl} | X_{rl}) \times \int p(Y_{ll} | X_{ll}) \times \int p(X_{rl}, X_{ll} | X_{lower}) dX_{lower} dX_{ll} dX_{rl} \quad (1)$$



(a) The Human Skeleton Model

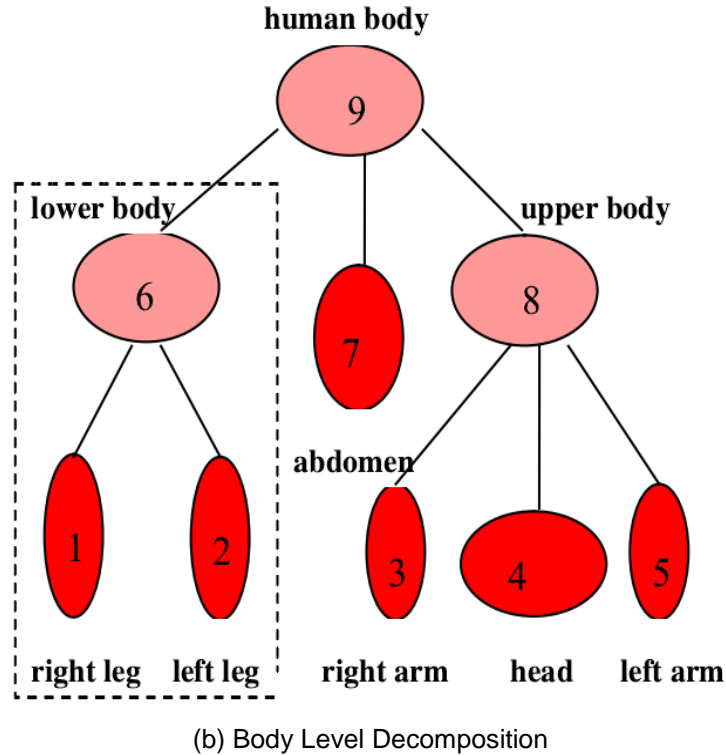


Figure 1. The Human Skeleton Model and its Decomposition

By formula 1 posterior probability to calculate the corresponding latent variable, get formula 2

$$\log p(X_{rl}, X_{ll}, X_{lower} | Y_{rl}, Y_{ll}) = \log p(Y_{rl} | X_{rl}) + \log p(Y_{ll} | X_{ll}) + \log p(X_{rl}, X_{ll} | X_{lower}) \quad (2)$$

In the paper, the kernel function of employed HGPLVM adopts the radius base function kernel which consists with formula 1. The whole hierarchical latent variable space is optimized in the following steps:

- (1) Initialize with PCA latent variables in low-dimension space to which all leaf nodes correspond;
- (2) For every single non-leaf node in tree-like model, in the space constituted by latent variables of all child nodes owned by one non-leaf node, use PCA to initialize the latent variable in low-dimension space to which the non-leaf node relates;
- (3) Take equation 1 and 2 to jointly optimize parameters and latent variables of the whole hierarchical model;

We select from CMU MoCap human body behavior database some representative subtle behaviors to form single person behavior database for our experiment. Also we create all related hierarchical latent variable space to those behaviors. It is shown in figure2. Based on the evaluation of the mean energy reserving percentage by PCA algorithm during HGPLVM initializing latent variables, we choose the dimension of each sub-manifold space. In the experiment, when the dimension of sub-manifold space is 2, the mean value is 96.2% of energy retaining ratio; when it's 3, the mean value is 97.9%. For the consideration of visualization and computational efficiency, we make 2 the dimension of each sub-manifold space.

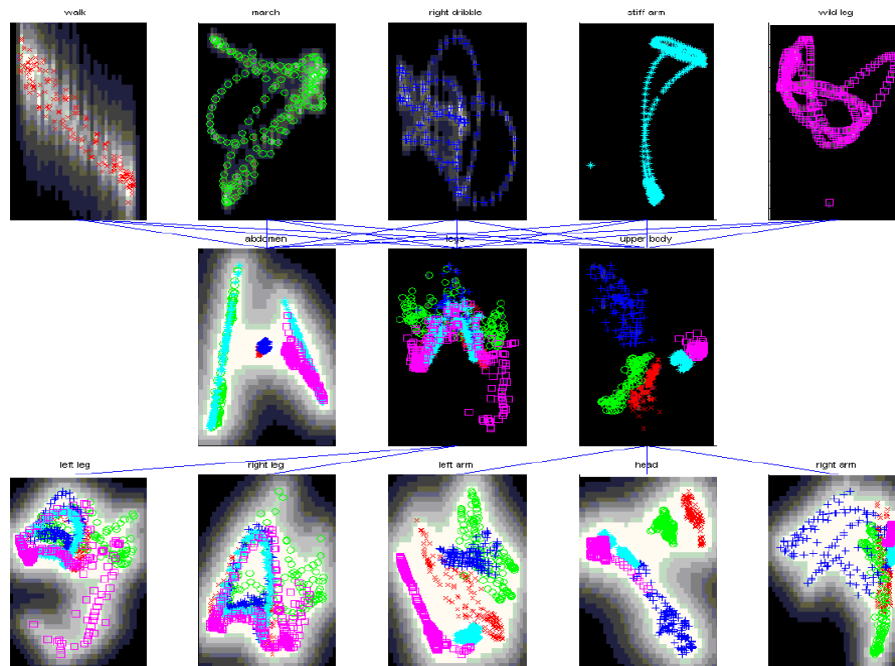


Figure 2. Hierarchical Latent Variable Space of Human Motion

2.2. Visualization of Partial Body Movement Trails

Since human behavior sequence has different length of sequence, to discover the common attributes of human body and each partial body movement trails, we normalize the length of all normalized behavior sequences in database in the range $[0, 1]$, *i.e.*, transforming all human behaviors to standardized behavior sequences without length attribute. So the instance of one man's behaviors can be regarded as the collection of normalized behavior sequences of different human body parts. Figure 3 shows separately some normalized behavior sequences of three classical low-dimension latent variable spaces. We see that normalized behavior sequences belonging to the same act are very similar. In Figure 3 (a), walking, walking with still hands and hyperbolic walking are in consistent trajectory in the latent variable space of right leg movements; also in other latent variable spaces, they show similar tracks. Hence, we think in the low-dimension space of movements of every body part, there must be some shared motions of different behavior types. The above normalized trails can be used to fetch such shared motions.

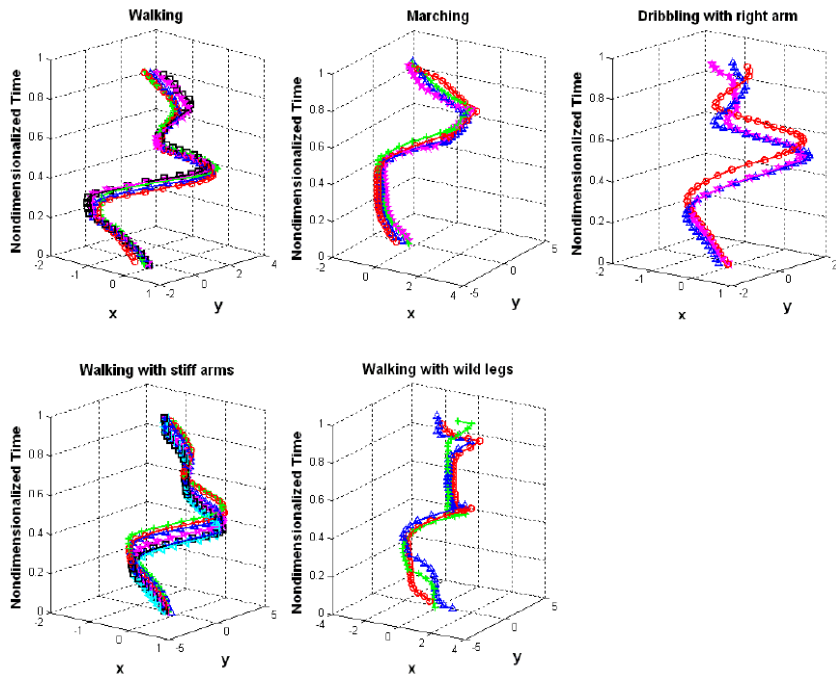
2.3. Trajectory Clustering

Trajectory clustering is an effective technique grouping similar tracks in trail collections. In the experiment, the dimension of latent variable spaces of all human part movements is 2. So the normalized behavior sequence (or normalized trail) of each human part movements can be expressed as $\{(a_1^x, a_1^y), (a_2^x, a_2^y), \dots, (a_N^x, a_N^y)\}$.

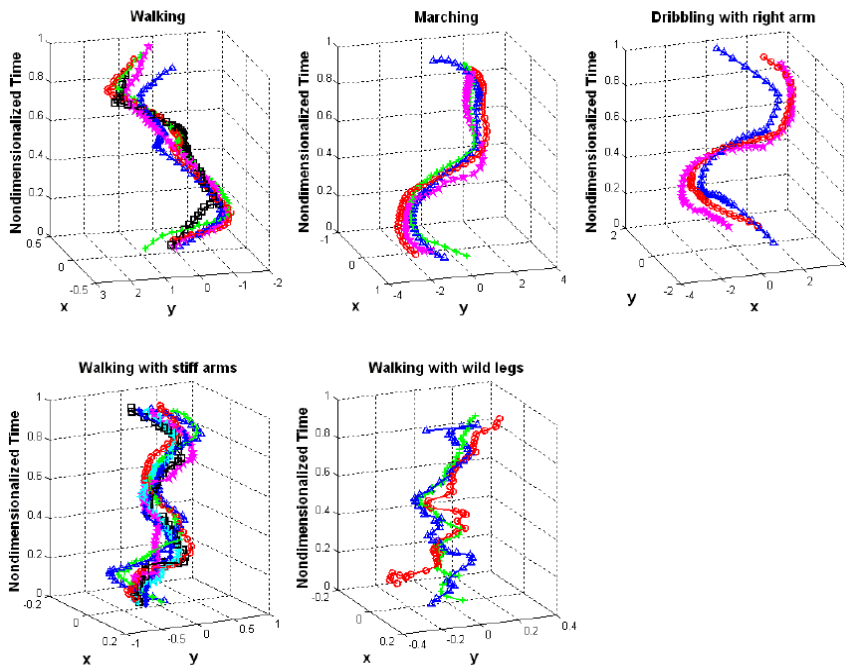
Regarding the clustering algorithm based on similarity, one key question is how to measure similarity of trails. Zhang *et. al.*, [12] compared the effect and efficiency of six similarity measurements in trajectory clustering, for motional tracks in outdoor monitoring scenes. Their experiments indicated that PCA + Euclidean distance acquired good clustering results, with high computational efficiency. Therefore here we adopt K-means clustering method of PCA + Euclidean distance to extract the motion pattern of each body part and parameterize movement models.

Take 50 sample points in each normalized behavior trajectory. That is $N=50$. As per the method in literature [13], we arrange in series the coordinate x and y of those sample

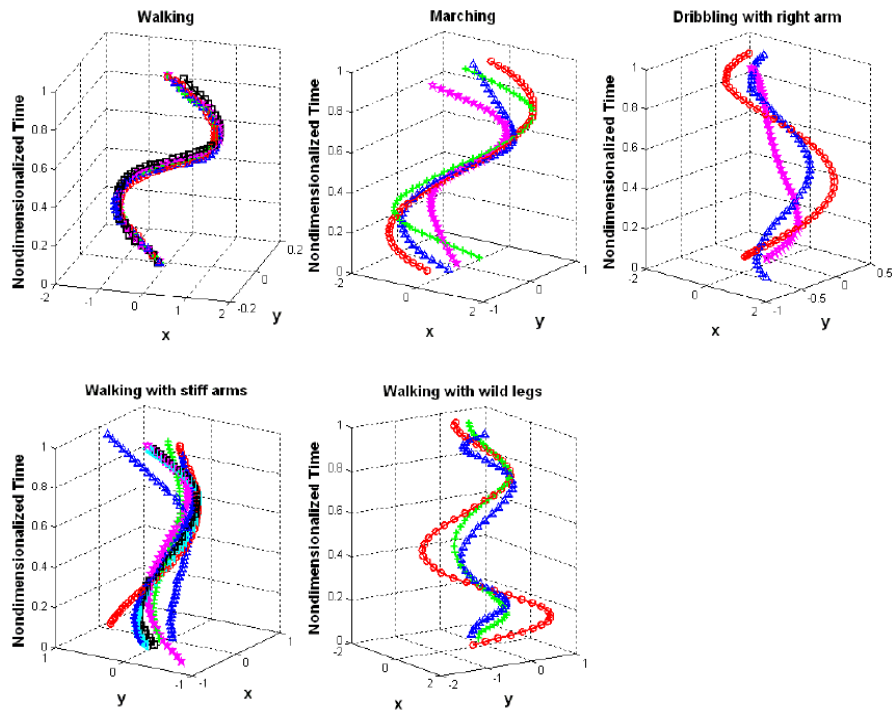
points. Figure 4 presents the trace of serial coordinates after normalization in two typical latent variable spaces. Obviously, after coordinate series, those trails are discernible as well. Figure 5 gives average tracks of five single movements in two typical latent variable spaces.



(a) Some instances of behavior sequence of low dimensional latent variable space motion in the right leg

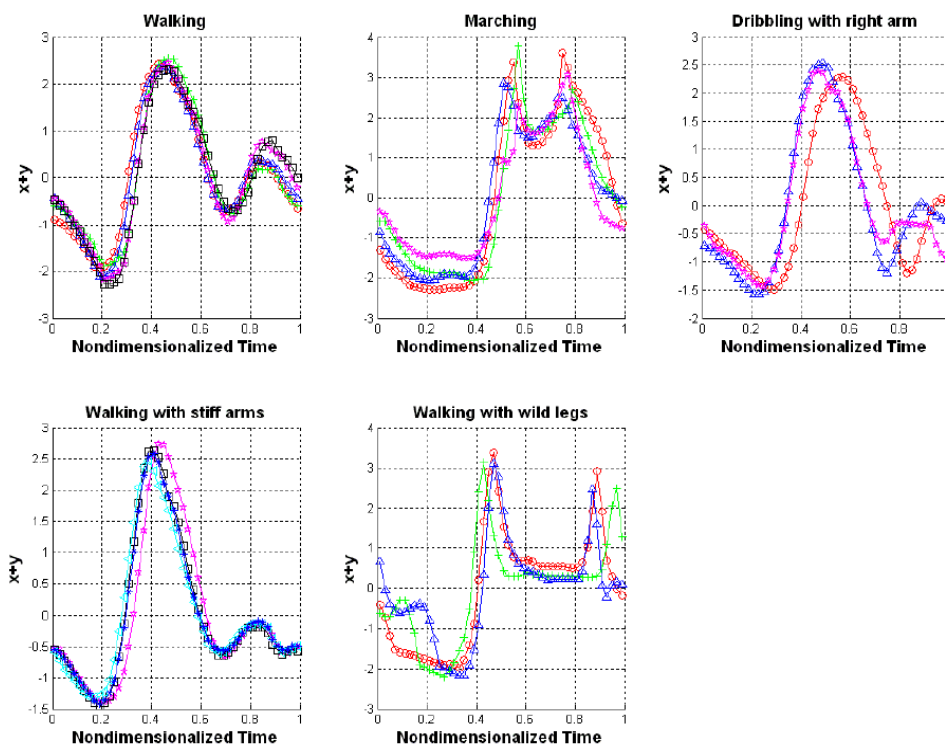


(b) Some instances of behavior sequence of low dimensional latent variable space movement in the right arm

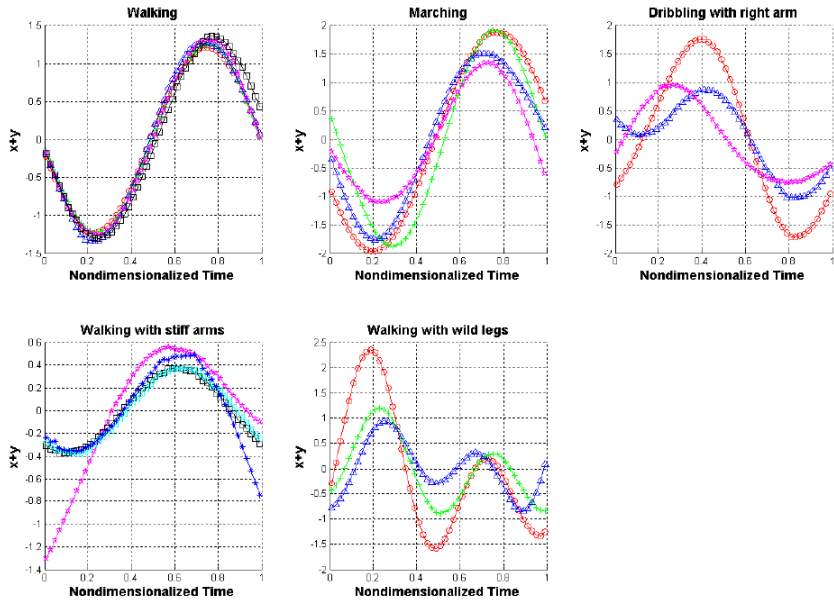


(c) Some instances of behavior sequence of top-level latent variable space

Figure 3. The Normalized Behavior Sequences of Three Typical Instances of Body Parts in the Latent Variable Space



(a) 1-D instances of low dimensional normalized trajectory latent variable space motion in the right leg

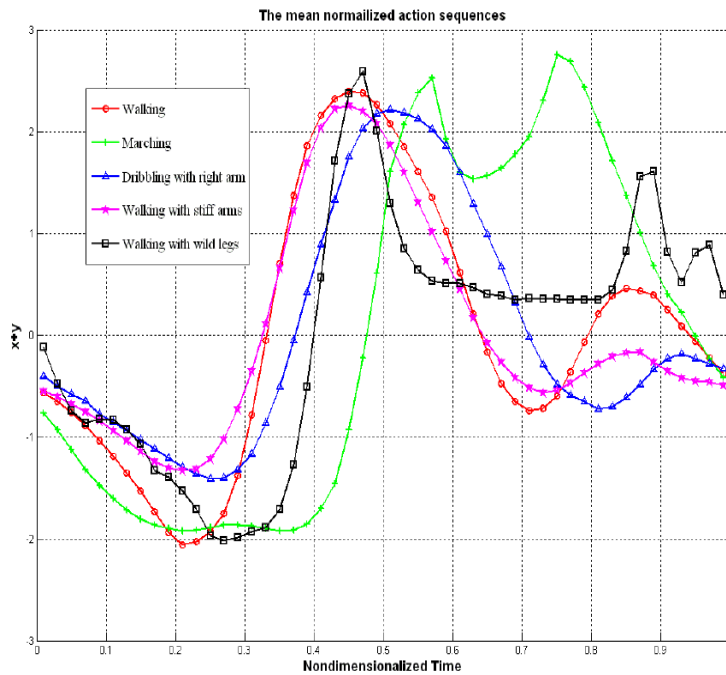


(b) 1-D instances of normalized trajectory latent variable space in the top case

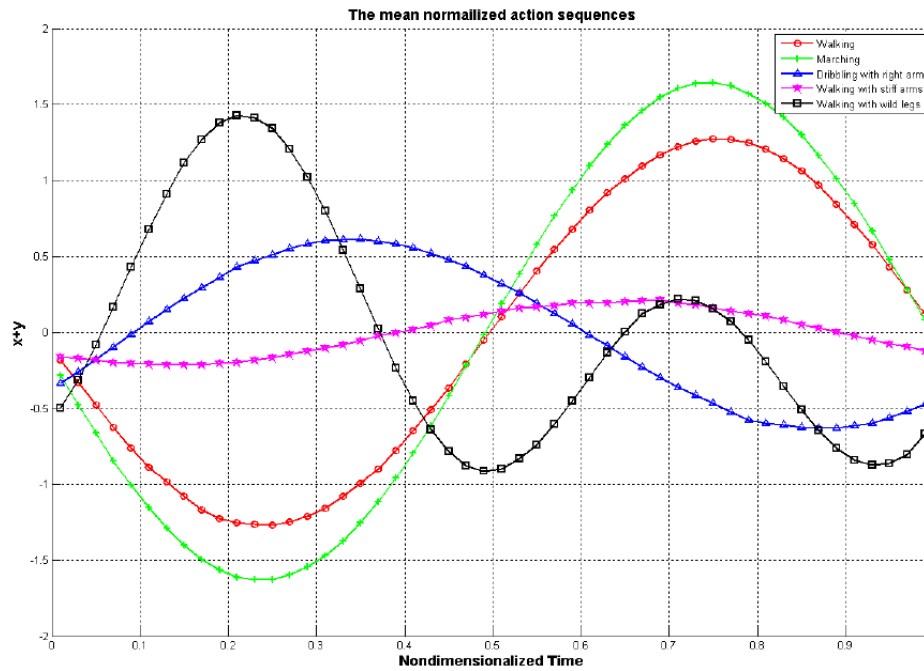
Figure 4. 1-D Instances of Normalized Trajectory Two Typical Latent Variables in Space

Eventually, normalized behavior trajectories can be shifted to PCA coefficient after coordinate x and y put in series. Assume A and B two parameterized trajectories after standardization. By estimating the Euclidean distance between PCA coefficients, we can measure their similar degree:

$$D(A, B) = \left[\sum_{k=1}^K (a_k^z - b_k^z)^2 \right]^{\frac{1}{2}} \quad (3)$$



(a) The average normalized trajectory of five latent in the right leg variable space



(b)The average corresponding normalized trajectory of five latent in the top of the variable space

Figure 5. The Average Normalized Trajectory of Five Kinds of Behavior of Two Typical Low Dimensional Latent in the Variable Space

The number of clustering center in K-means algorithm is made empirical value. Meanwhile we utilize correct clustering rate (CCR) to measure the efficiency of K-means clustering. In clustering, the initial clustering center chooses in a random way, repeating five times the clustering process of each human body movement in latent variable space, and considers the average value of five CCRs as the final result. We take the mean value after five clustering as extracted motion pattern. So each movement mode is expressed as one K-dimension vector. It is shown in Figure 6.

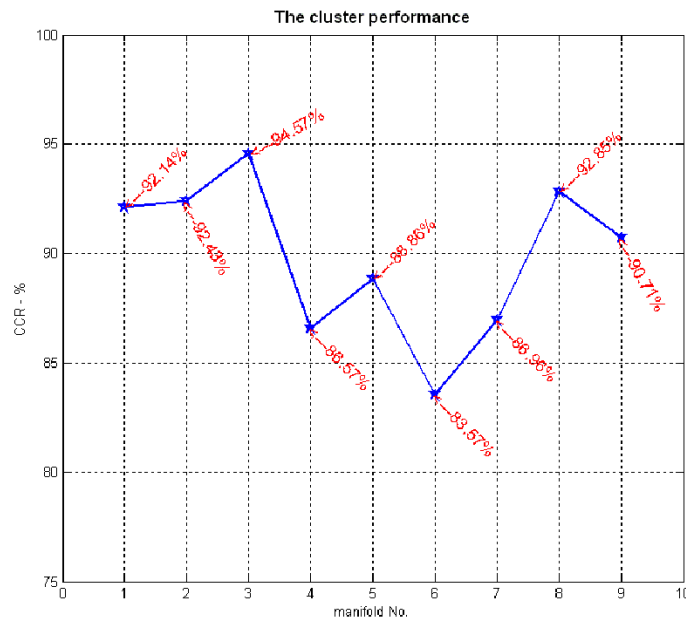


Figure 6. Clustering Correct Rate in Nine Body Parts Latent Variable Space

3. Single Behavior Modeling based on Discriminative Models

Man's behavior recognition can be looked as where the observation vector in each frame predicts behavior tag. Based on the tree-like structure of hierarchical latent variable space of human movements built in 2.1, we make one cascaded conditional random field (CRF) model which has the identical topological structure. With the model, we model the relationship between movements of each human body part and the related motional patterns. In the model, sub CRF relevant to each node has its own observation vector. Leaf node relevant sub CRF models' input feature, *i.e.*, observation vector, is only the coordinate value of latent variables in the latent variable space. Input features of non-leaf nodes' corresponding sub CRF model include observation vectors in the related latent variable space as well as the input of its child nodes' relative CRF model. All sub CRF models' inputs are motional pattern sequence of relevant body parts. So behaviors of the whole body can be represented as the motion mode collect of various body parts.

Each sub CRF is expressed by a graph model having a state chain. Suppose s is the motion model tag of one motion model collection L , observation vector o , observation sequence S , motion mode tag sequence T and cluster C in graph model G . Then CRF models directly the conditional probability of the given observed state sequence like:

$$P_{\theta}(S|O) = \frac{1}{Z(O)} \prod_{c \in C} \Phi(S_c, O_c) \quad (4)$$

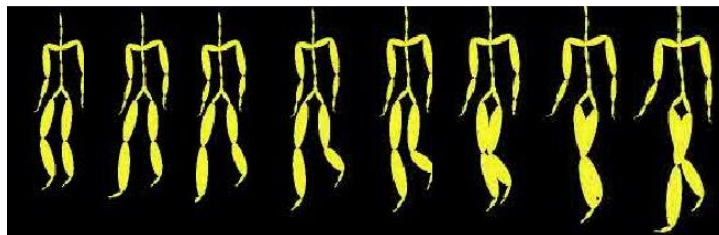
$$Z(O) = \sum_s \prod_{c \in C} \Phi(S_c, O_c)$$

4. Experiment Design and Discussion

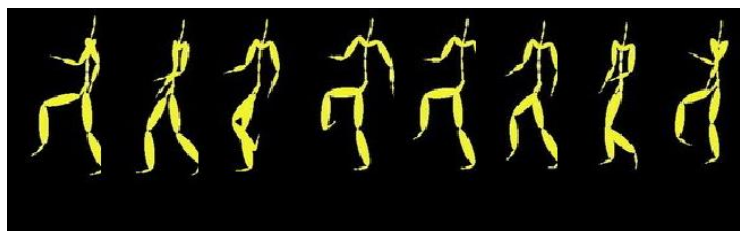
We used motion capture data to validate the proposed single behavior recognition method. The extraction of motion mode and training of discriminative model were both directly tested with motion capture data. In the experiment, we also evaluated the influences of different human movements on behavior recognition of the whole human body.

4.1 Single behavior database

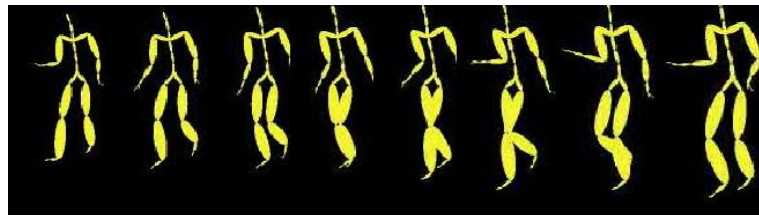
As seen from Figure 7, we chose five behaviors from CMU MoCap database to create our single behavior database. The five behaviors refer to walking, goose-step, dribbling with right hand, walking with still hands and hyperbolic walking, which are different with one or more body movements.



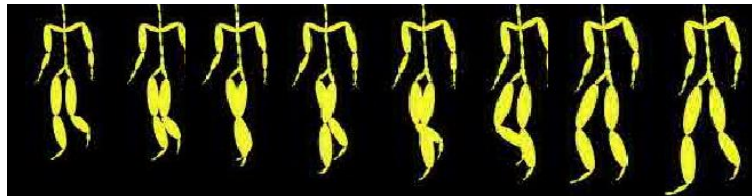
(a) walking



(b) goose-step



(c) dribbling with right hand



(d) walking with still hands



(e) hands and hyperbolic walking

Figure 7. Five Kinds of Single Behavior in CMU MoCap Database

We separated intentionally behavior sequences chosen from MoCap behavior database into 157 behavior sections, each including one behavioral cycle of some behavior. See details in Table 1.

Table 1. Fine Human Behavior Library

	number of behavioral segments	average segments number of behavior
walking	69	135
goose-step	23	146
dribbling with right hand	11	124
walking with still hands	26	160
hands and hyperbolic walking	28	129

To discriminate training set and testing set, we divided the whole behavior database to three sections, respectively, dataset 1, dataset 2 and dataset 3 (See Table 2). With cross-validation method, we tested the algorithm here.

Table 2. Data Set Division

Fine behavior	data set1	data set2	data set3
walking	1-23	24-46	47-69
goose-step	1-8	9-16	17-23
dribbling with right hand	1-4	5-8	9-11
walking with still hands	1-9	10-18	19-26
hands and hyperbolic walking	1-9	10-18	19-28

4.2. Results of Experiments with Motion Capture Data

Figure 8 graphs the recognition rate of cascade conditional field identifying motion models in latent variable space of each human body behavior. Table 3 provides confusion matrix of recognition of five single behaviors. We assessed when the window width is respectively $w=0$ and $w=3$, the cognitive results of CRF model. When $w=1$, the recognition performance was obviously better than $w=0$. Through plentiful tests, we also found when $w=1$, the recognition result is satisfying. That's why no recognition rate was provided for bigger value of w .

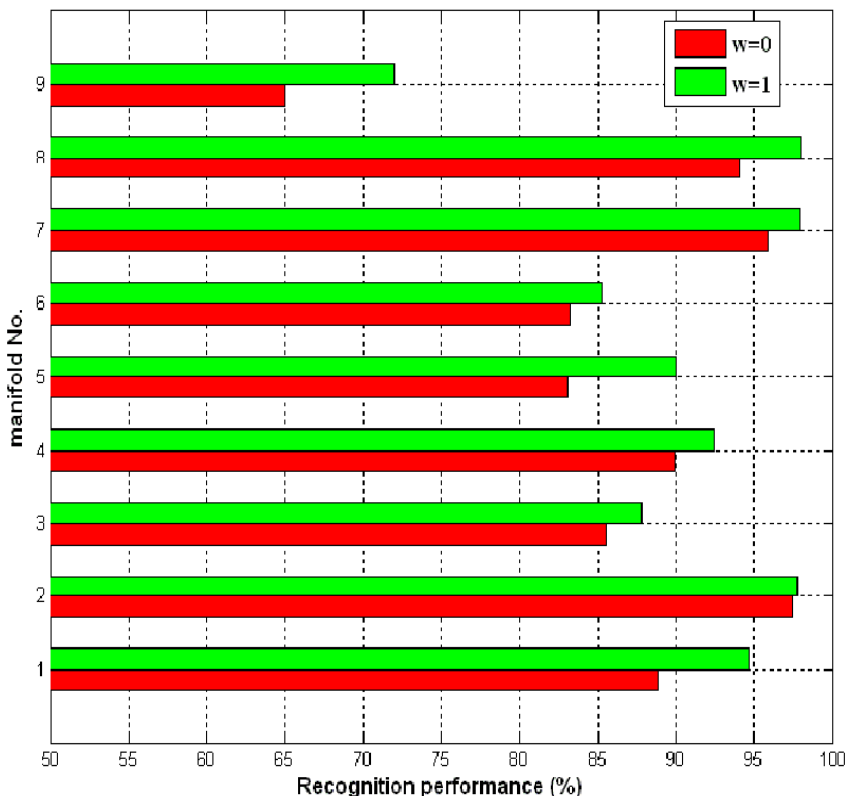


Figure 8. The Recognition Rate of Stacked CRF Movement Pattern in the Sports Body Latent Variable Space

Table 3. The Identification Result based on Motion Capture Database

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	96.1	0.87	2.63	0	0.45
goose-step	1.35	98.65	0	0	0
dribbling with right hand	0.76	0	98.98	0	0.36
walking with still hands	0.97	0	4.17	92.7	2.18
hands and hyperbolic walking	0	0	0.08	4.79	95.13

(a) $w=0$

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	98.28	1.75	0	0	0
goose-step	0	100	0	0	0
dribbling with right hand	0	0	100	0	0
walking with still hands	0	0	1.87	96.22	1.69
hands and hyperbolic walking	0	0	0	3.14	98.56

(b) $w=1$

To estimate the impact of different body part movements on the whole body behavior recognition, we artificially overlooked some parts and evaluated accordingly the recognition result of single behavior. Table 4 shows confusion matrix of recognition when several human body parts were ignored. From it, we learnt that when movements in the whole lower part like left & right legs and lower part were neglected, goose-step and hyperbolic walking were pretty distinguishable. That's why they were not correctly identified. On the contrary, since five behaviors were not much discernible in the head part, when head movements were removed, recognition of all behaviors was basically none affected.

Table 4. The Removal of Some Body Parts when the Rate of Recognition

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	100	0	0	0	0
goose-step	100	0	0	0	0
dribbling with right hand	0	0	100	0	0
walking with still hands	0	0	1.87	98.59	0
hands and hyperbolic walking	0	0	0.03	99.96	0

(a) Ignore the right leg, left leg and the lower part of the body

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	98.55	0.45	0	0	0
goose-step	40.25	59.69	0	0	0
dribbling with right hand	100	0	0	0	0
walking with still hands	100	0	0	0	0
hands and hyperbolic walking	3.69	0.07	0	0	95.56

(b) Ignore the right arm, left arm and upper body

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	98.55	0.75	0	0	0
goose-step	40.04	59.96	0	0	0
dribbling with right hand	0	0	100	0	0
walking with still hands	0	0	1.81	96.22	1.98
hands and hyperbolic walking	0	0	0.02	3.36	98.62

(c) Ignore the left leg

Human behavior	walking	goose-step	dribbling with right hand	walking with still hands	hands and hyperbolic walking
walking	98.28	1.72	0	0	0
goose-step	0	100	0	0	0
dribbling with right hand	4.07	0	0	76.41	19.56
walking with still hands	0	0	0	96.03	1.99
hands and hyperbolic walking	0	0	0	3.15	98.85

(d) Ignore the right arm

5. Conclusion

With the prior knowledge about human body decomposition of physiology, we established hierarchical latent variable space of human movements and in it we analyzed the common features of various human body part movements, *i.e.*, motion modes. With bottom-top discriminative behavior modeling frame, we created the model of the whole human body and therefrom we developed one cascade CRF model to recognize motion patterns of various body parts. Experiments with CMU capture database proved its effectiveness and robustness.

References

- [1] L. Wenyang, "Study on behavior recognition algorithm based on intelligent mobile phone sensor", Jilin University, (2014).
- [2] S. Aimeng, "The recognition method of single space-time interest points and behavior based on interactive behavior", Jilin University, (2014).
- [3] X. Yajing, "Two value of local space-time behavior recognition method based on patterns", Jilin Agricultural University, (2013).
- [4] G. Ping, "Human behavior analysis based on video", Beijing Jiaotong University, (2012).
- [5] F. chaoxia, "Study on human target detection and behavior recognition based on video stream", North Central University, (2013).
- [6] Yang Jiahui. Human behavior recognition based on spatio-temporal interest point. Nanjing University of Posts and Telecommunications, 2014
- [7] Wu Songlin. Many people sitting behavior recognition and analysis method research based on. Yanbian University, 2010

- [8] Deng sweet. Human behavior recognition research based on hierarchical model. University of Science & Technology China, 2010
- [9] Li Junfeng. Analysis of people visual interaction based on. Beijing Institute of Technology, 2010
- [10] Cheng xun, Wu Zhijie. Human behavior recognition method based on key frame. Computer engineering and applications, 2013,18:134-137.
- [11] Wang Jiangtao. Study on video object detection, tracking and behavior recognition based on. Nanjing University of Science and Technology, 2008
- [12] Zhang Z., Huang K. And Tan T. Comparison of similarity measure for trajectory clustering in outdoor surveillance scenes. IEEE International Conference on Pattern Recognition, 2006.
- [13] Bashir F.I., Khokhar A.A. and Schonfeld D. Segmented trajectory based indexing and retrieval of video data. IEEE International Conference on Image Processing, 2003.

Author



Yanhua Chen, she received her B.S degree from Sichuan Normal University. She received her M.S degree from Sichuan Normal University. She is a senior experimenter in Modern education technology center of Neijiang Normal University. Her research interests include Computer Application.