

## An Improved Hidden Markov Model based on Human Motion Detecting Approach via Multi-View Image Sequences

Yuhua QIU, Jing HUANG, Weiwei DING

*Department of Sports*  
China University of Geosciences (Wuhan)  
430074 China

**Abstract** — In this paper, we concentrate on the problem of human motion detecting from multi-view image sequences based on a modified hidden Markov model, and human motion detecting is of great importance in computer vision. Firstly, the structure of human skeleton model is given, which refers to a local coordinate system. In this model, the human bones follow the Parent-Child relationship. In particular, the upper node of human skeleton model is named skeleton root, and it can connect with the spine root. Apart from the spine root, the remainder parts of the human body can be extended from the leg, including: 1) thigh, 2) shin, 3) foot and 4) toes. Secondly, we locate the object of human body from multi-view image sequences, and then propose a novel method to find human's head and shoulder. Thirdly, as a complete human motion representation may be the set of all three-dimension points on a specific actor, in this paper, we represent human motion as four-dimension points in real environment. Next, a modified Hidden Markov model is designed to represent a symbol sequence, and then human motions can be obtained from the outputs of the proposed Hidden Markov model. To demonstrate the performance of the proposed method, a series of experiments are conducted. Experimental results show that compared with other existing methods, the proposed approach can effectively enhance the accuracy of human motion detection.

**Keywords** -- *Hidden Markov model; Human motion detecting; Multi-view image sequences; Stationary probability distribution; Observation sequence*

### I. INTRODUCTION

As a new research field in visual computer domain, human action recognition refers to a multi-interdisciplinary research subject, which is interested by the researchers in the fields of image/video processing, computer vision, pattern recognition, statistic learning, and artificial intelligence<sup>[1][2]</sup>. This topic can analyze the content of the raw data (videos, images), extract valuable cues which are discriminative about to human actions, and then construct the relationship between the original input data and the high level semantic information. Developing the method of human motion detecting can provide chance for many applications, such as intelligent video surveillance, human-computer interaction, virtual reality, and so on. Considering the wide application prospects and the high theory meaning, in recent years, there are a lot of researches in computer vision domain<sup>[3-5]</sup>.

With the rapid development of human motion detecting method, the efficient and rapid detection of human body moving information has become into reality. Human motion detecting has been widely exploited in many fields, for example sports competition, animated games, film productions and so on<sup>[6][7]</sup>. Currently, the human motion detecting algorithm has been classified to 4 systems, that is, 1) Mechanical Motion Capture system, 2) Acoustic Motion Capture system, 3) Magnetic Motion detection system and 4) Optical Human Motion Capture system<sup>[8][9]</sup>. Human motion detection based on motion detection device is belonged to

one of important problems in virtual reality system. As is well known that human body movement data can be detected via human motion capture devices. Furthermore, this study can be exploited to human-computer interaction, education, medical research, film animation and other applications<sup>[10][11]</sup>. As is shown in Fig.1, some examples for human motion detection are provided, and we can find that several key points must be determined to represent the human body.

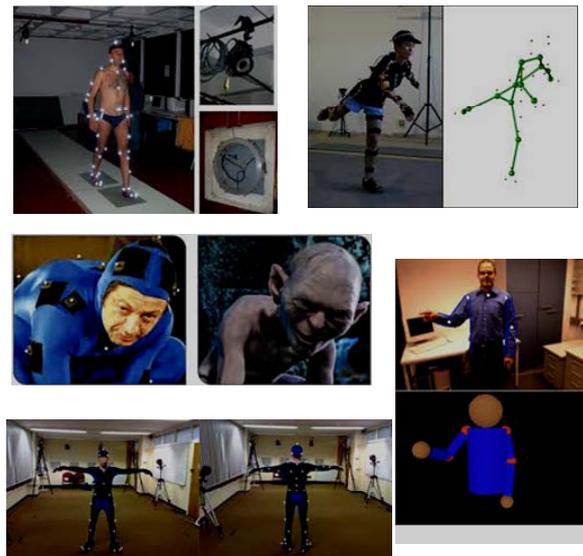


Fig. 1 Examples for human motion detection

Based on the above analysis, in this paper, we focus on the problem of detecting human motions from multi-view image sequences. The rest of the paper is organized as follows. Section 2 illustrates related works about human motion detection. In section 3, pre-processing of the human motion detection system is provided. Section 4 presents experimental results and provides related analysis. Finally, the conclusions are drawn in section 5.

II. RELATED WORKS

Human motion detection has been a hot research field in recent years, and several approaches have been proposed. In the following section, we will survey on the related works of our works.

Zhang et al. proposed an approach for keyframe extraction from human motion capture data using a multiple population genetic algorithm. Particularly, in this paper, the fitness function is defined to follow the goals of minimal reconstruction errors and the optimal compression rate, in which multiple initial populations are subjected to co-evolution. Furthermore, the multiple population genetic algorithm utilizes global and local search<sup>[12]</sup>.

Royo et al. proposed a novel calibration procedure for optical human motion capture systems. Firstly, initial estimators of intrinsic and extrinsic parameters are sought. Secondly, a simultaneous nonlinear optimization of all parameters is performed to identify the optimal values, which minimize the objective function. The objective function can minimize the above two errors<sup>[13]</sup>.

Jiang et al. proposed a novel approach utilizing singular value decomposition for extracting embodied knowledge from the time-series data of the motion. In this paper, the authors compose a matrix from the time-series data and exploit the left singular vectors of the matrix as the patterns of the motion and the singular values as a scalar. Particularly,

this paper can get a higher correct categorization ratio than principal component analysis and correlation efficiency<sup>[14]</sup>.

Guan et al. developed a compressive sensing based method to classify human motions based on pyroelectric infrared sensors. In this paper, the authors represented a human motion as a spatio-temporal energy sequence and then discover it from an infrared radiation domain. In order to obtain this sequence, a mask should be utilized to separate the object space to small meshes. 360 compressive STESs of ten aerobic exercises performed by six persons are used as the experimental dataset for performance evaluation<sup>[15]</sup>.

Afterwards, as this paper utilizes the Markov model to solve the human motion detection problem, and in the following parts, we will discuss the applications of Markov model in many different fields.

Moghaddass et al. utilized the nonhomogeneous semi-Markov model to forecast the remaining useful life in minimizing the overall maintenance cost of mechanical systems<sup>[16]</sup>. Fort et al. exploited Hidden Markov Models to estimate life parameters<sup>[17]</sup>, Qiao et al. used Hidden Markov Model to classify dynamic texture<sup>[18]</sup>, Manandhar et al. developed a Multiple-Instance Hidden Markov Model to detect GPR-Based Landmine<sup>[19]</sup>, Lin et al. used Markov model to analyze economic impact of psychiatric relapse and recidivism among adults<sup>[20]</sup>, Durmaz et al. used Markov model to implement polymer assembly<sup>[21]</sup>, Gong et al. utilized Markov model to dynamically simulate vegetation abundance in a reservoir riparian zone<sup>[22]</sup>, and Gahrooei et al. exploited Hidden Markov Models to infer traffic signal phases from turning movement counters<sup>[23]</sup>.

III. PRE-PROCESSING OF THE HUMAN MOTION DETECTION SYSTEM

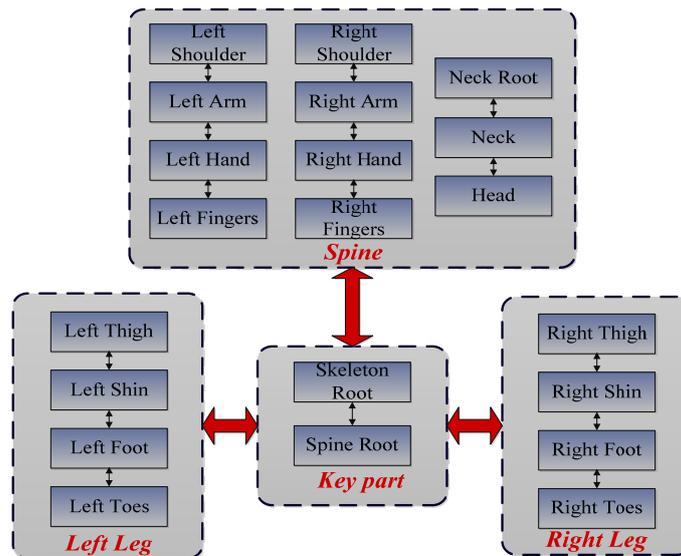


Fig.2 Structure of human skeleton model.

In this section, we illustrate the pre-processing steps of the human motion detection system. Firstly, the structure of human skeleton model is described in Fig.2.

3D human skeleton model denote a local coordinate system, in which the bones follow the Parent-Child relationship. The upper node of human skeleton model is called skeleton root, and it is able to connect with the spine root. Additionally, the spine root has the ability to connect legs with spine. Apart from the above parts, the rest parts of the human body are extended from the leg, which are 1) thigh, 2) shin, 3) foot and 4) toes. Among all the parts of human body, spine is of great important part, because it can connect shoulder, arm, hand, neck with head.

Secondly, before detecting the human motions, we should illustrate how implement the pre-process steps (shown in Fig. 3).

We suppose that  $B_0(x, y)$  refers to the initial background image that is a frame image sequences, and symbol  $I(x, y)$  means the current frame image. Moreover,  $B_{k-1}(x, y)$  denotes the background image of the last moment, and  $B_k(x, y)$  refers to the current background image, and it is defined as follows.

$$B_k(x, y) = (1 - \alpha) \cdot B_{k-1}(x, y) + \alpha \cdot I(x, y) \quad (1)$$

We utilize the current image sequence and the background frame, and then use the absolute value to obtain the foreground object image  $F_k(x, y)$ .

$$F_k(x, y) = I(x, y) - B_k(x, y) \quad (2)$$

Assuming that the symbol  $T$  means the adaptive threshold, and  $U_k(x, y)$  is the binary foreground object image.

$$U_k(x, y) = \begin{cases} 255, & F_k(x, y) < T \\ 0, & otherwise \end{cases} \quad (3)$$

Afterwards, we will explain how to find human's head and shoulder in multi-view image sequences. Assuming that  $A$  means the number of pixels of connected region,  $(X_1, Y_1)$  refers to the upper left coordinate of the rectangle frame, and  $(X_2, Y_2)$  is the lower right coordinate of the rectangle frame. Based on the above analysis, the concrete equation is represented as follows.

$$X_g = \sum \frac{X_i}{A}, Y_g = \sum \frac{Y_i}{A} \quad (4)$$

$$X_1 = \text{Min}(X_i), Y_1 = \text{Min}(Y_i) \quad (5)$$

$$X_2 = \text{Min}(X_i), Y_2 = \text{Min}(Y_i) \quad (6)$$

Then, the barycentric distance between human's head and shoulder is defined as follows.

$$D = \sqrt{(X_{g1} - X_{g2})^2 + (Y_{g1} - Y_{g2})^2} \quad (7)$$

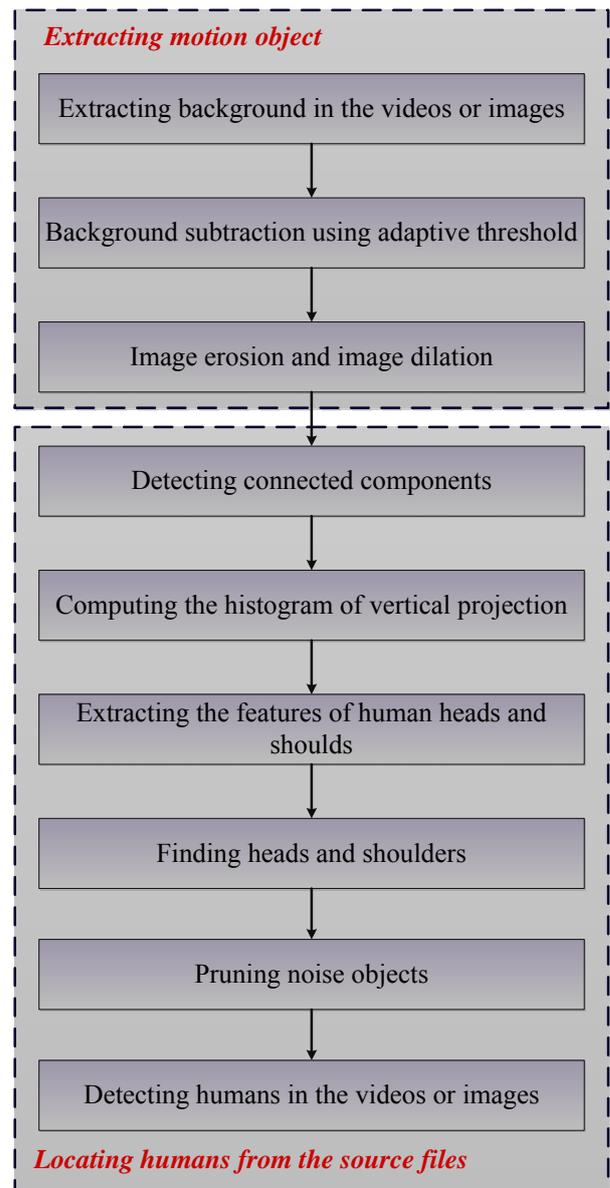


Fig. 3 Flowchart of the pre-processing steps in the human motion detection system

IV. THE PROPOSED ALGORITHM

Human action refers to the movement of humans to execute a task in a short period of time. The action may be simple or complex relying on the number of body limbs. We suppose that a complete human action representation may be the set of all three-dimension points on a specific actor. Hence, we can represent human actions as four-dimension points in real environment as follows.

$$A_{4D} = \begin{pmatrix} X_j^{T_1} & X_j^{T_2} & \dots & X_j^{T_p} \\ Y_j^{T_1} & Y_j^{T_2} & \dots & Y_j^{T_p} \\ Z_j^{T_1} & Z_j^{T_2} & \dots & Z_j^{T_p} \end{pmatrix} \quad (8)$$

where  $X$ ,  $Y$ , and  $Z$  represent the state-space representation of a point in the four-dimension plane of a human performing a specific motion.

In order to extend the Markov concept, we revise the standard Markov models by adding a hidden layer which permits us to consider the evolution of the state transition. The formal description of the hidden Markov model can be represented as  $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ . Moreover, for a specific state  $S_j$ , function  $\hat{b}_j()$  is defined using the space of the observation vectors  $\Omega$ . For a specific time slot  $s$  ( $s \in [1, T]$ ), fuzzy densities is defined as follows.

$$\hat{b}_1(Q_s), \hat{b}_2(Q_s), \dots, \hat{b}_M(Q_s) \quad (9)$$

where  $M$  means a fuzzy set on the set which contains  $M$  different states. Therefore,  $T$  different fuzzy sets on the set of  $M$  states are constructed. Hence,  $\hat{b}_k(Q_s)$  refers to a membership value of  $Q_s$  in the state  $S_k$ .

Next, the set of  $K$  training observation sequences is given by the following equation.

$$Q = \{Q^1, Q^2, \dots, Q^K\} \quad (10)$$

where  $Q^k = (Q_1^k, Q_2^k, \dots, Q_{T_k}^k)$  denotes the  $k$ -th observation sequence. Afterwards, the parameters of the hidden Markov model can be computed by solve the following optimization problem:

$$\hat{P}(Q|\hat{\lambda}) = \prod_{k=1}^K \hat{P}(Q^k|\hat{\lambda}) = \prod_{k=1}^K \hat{P}^k \quad (11)$$

Hence, we can allocate to each state a possible outcome symbol to quantify the level of each state's performance, and a modified hidden Markov model is proposed and this new model is named semi-Hidden Markov model. The

formal definition of our proposed semi- Hidden Markov model is described as follows.

Supposing that each new state is obtained by the sequence of  $z$ , and the lower the value of  $z$ , the lower of the number of states.  $P^* = \{p_1^*, p_2^*, \dots, p_m^*\}$  is the stationary probability distribution, the length of statistical inertia is represented as  $z \in \mathbb{N}, z \geq 1$ , the weight of statistical inertia  $w \in \mathfrak{R}, 0 \leq w \leq 1$ . Additionally, the probability of varying the state  $\pi \in \mathfrak{R}, 0 \leq \pi \leq 1$ .

Afterwards, we utilize a semi-Hidden Markov model  $\lambda = \{P^*, z, w, \pi\}$  to represent a symbol sequence  $O = \{o_1, o_2, \dots, o_T\}$  by the following steps.

Step 1: Obtaining  $z$  random symbols which follow the probability distribution  $P = P^*$ , and we assume that  $F$  refers to a relative frequency distribution of the above  $z$  symbols

Step 2: Let the initial state by the following probability distribution:

$$P = \tau \cdot F + (1 - \tau) \cdot P^* \quad (12)$$

Step 3: Obtaining a symbol of the alphabet based on the probability distribution  $P$ , and then output to the proposed sequence.

Step 4: Producing a random real number  $\delta \in [0, 1]$ .

Step 5: If parameter  $\delta$  is larger than  $\pi$

Step 6: Then repeat the third step.

Step 7: Else jump to the fifth step.

Step 8: Calculating the probability distribution of the new state, and set the symbol  $F$  refers to the corresponding to the frequency distribution of  $z$ .

V. EXPERIMENT

To evaluate the effectiveness of our proposed algorithm, MPI08 Database[4] is chosen to make performance evaluation, and MPI08 is developed by multi-sensor fusion for 3D full-body human motion detection. MPI08 Database is made up of 5 parts, that is, 1) sequences: multi-view sequences obtained from eight calibrated cameras, 2) silhouettes: binary segmented images obtained with chroma-keying, 3) meshes: three dimension laser scans for each of the 4 actors in the dataset and the registered meshes with inserted skeleton, 4) projection matrices: one for each of the eight cameras, 5) Orientation data: raw and calibrated and sensor orientation data. Some images of the MPI08 Database are described in the following figure.

Based on the MPI08 Database, we compare the performance our algorithm with other human motion detection approaches, that is, 1) ECPF [5], 2) SVD[14], and 3) PIS[15]. Additionally, five kinds of motions are used in this experiments, including: 1) M1: Cartwheel, 2) M2: Jumping jack, 3) M3: Skiing, 4) M4: Rotating both arms and

4) M5: Kicking. Firstly, we will use the confusion matrix to test the performance of human motion detection, and the results of each method are shown in the following tables.



Fig. 4 Examples of the motion “Jumping jack and skiing” in MPI08 database.

TABLE. 1 CONFUSION MATRIX FOR ECPF

|    | M1   | M2   | M3   | M4   | M5   |
|----|------|------|------|------|------|
| M1 | 0.81 | 0.02 | 0.11 | 0.04 | 0.02 |
| M2 | 0.1  | 0.75 | 0.08 | 0.02 | 0.05 |
| M3 | 0.06 | 0.13 | 0.71 | 0.04 | 0.06 |
| M4 | 0.02 | 0.02 | 0.04 | 0.86 | 0.06 |
| M5 | 0.18 | 0.02 | 0.04 | 0.05 | 0.71 |

TABLE. 2 CONFUSION MATRIX FOR SVD

|    | M1   | M2   | M3   | M4   | M5   |
|----|------|------|------|------|------|
| M1 | 0.89 | 0.02 | 0.03 | 0.04 | 0.02 |
| M2 | 0.03 | 0.84 | 0.02 | 0.08 | 0.03 |
| M3 | 0.02 | 0.03 | 0.89 | 0.03 | 0.03 |
| M4 | 0.06 | 0.06 | 0.04 | 0.82 | 0.02 |
| M5 | 0.03 | 0.08 | 0.04 | 0.03 | 0.82 |

TABLE. 3 CONFUSION MATRIX FOR PIS

|    | M1   | M2   | M3   | M4   | M5   |
|----|------|------|------|------|------|
| M1 | 0.44 | 0.22 | 0.13 | 0.12 | 0.09 |
| M2 | 0.15 | 0.63 | 0.05 | 0.06 | 0.11 |
| M3 | 0.11 | 0.3  | 0.51 | 0.05 | 0.03 |
| M4 | 0.03 | 0.05 | 0.04 | 0.78 | 0.1  |
| M5 | 0.05 | 0.04 | 0.18 | 0.06 | 0.67 |

TABLE. 4 CONFUSION MATRIX FOR THE PROPOSED ALGORITHM

|    | M1   | M2   | M3   | M4   | M5   |
|----|------|------|------|------|------|
| M1 | 0.96 | 0.01 | 0.02 | 0.01 | 0    |
| M2 | 0    | 0.99 | 0.01 | 0    | 0    |
| M3 | 0.03 | 0.02 | 0.93 | 0.01 | 0.01 |
| M4 | 0.02 | 0.01 | 0.01 | 0.94 | 0.02 |
| M5 | 0.01 | 0    | 0    | 0    | 0.99 |

Next, we integrate all the above experiments, and the human motion detection accuracy for these approaches are illustrated in Fig. 5.

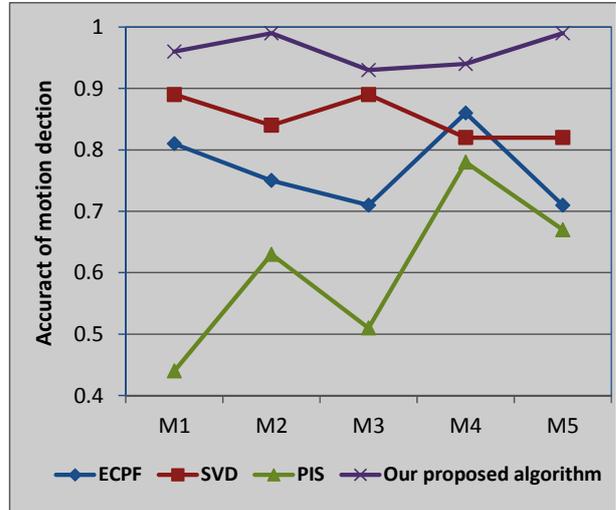


Fig. 5 Performance of motion detection accuracy for different approaches.

In the following section, we will show the recognition accuracy by the proposed algorithm under each single view image and multi-view images (shown in Table.1). Integrating all the above experiments, the conclusions can be drawn that our proposed can achieve higher human motion detection accuracy than other methods for the five different gestures.

## VI. CONCLUSION

This paper focuses on the problem of human motion detecting from multi-view image sequences using a modified hidden Markov model. In order to locate the object of human body from multi-view image sequences, we propose a new approach to discover human’s head and shoulder. The main innovations of this paper lie in that we provide an improved Hidden Markov model to represent a symbol sequence, and then human motions can be extracted from the outputs of the proposed Hidden Markov model. Finally, Experimental results demonstrate the effectiveness of our proposed algorithm.

## REFERENCES

- [1] Armon Leah, Caplan S. Roy, Eisenbach Michael, Friedrich Benjamin M., “Testing Human Sperm Chemotaxis: How to Detect Biased Motion in Population Assays”, PLOS ONE, vol. 41 , No. 4, pp. 657-680, 2012
- [2] Kim Moon Hwan, Park Jin Bae, Ra In Ho, Joo Young Hoon, “Hybrid silhouette extraction method for detecting and tracking the human motion”, Advances in Natural Computation, vol. 41 , No. 4, pp. 687-695,2006
- [3] Xiao Jun, Feng Yinfu, Ji Mingming, Yang Xiaosong, Zhang Jian J., Zhuang Yueting, “Sparse motion bases selection for human motion denoising”, Signal Processing, vol. 110 , No. 4, pp. 657-680,2015.

- [4] Andreas Baak, Thomas Helten, Meinard Muller, Gerard Pons-Moll, Bodo Rosenhahn, Hans-Peter Seidel, "Analyzing and Evaluating Markerless Motion Tracking Using Inertial Sensors", European Conference on Computer Vision (ECCV Workshops, vol. 41 , No. 4, pp. 139-152,2010).
- [5] Liu Jigang, Liu Dongquan, Dauwels Justin, Seah Hock Soon, "3D Human motion tracking by exemplar-based conditional particle filter", Signal Processing, vol. 41 , No. 4, pp. 164-177 ,2015.
- [6] Lim Chern Hong, Vats Ekta, Chan Chee Seng, "Fuzzy human motion analysis: A review", Pattern Recognition, vol. 41 , No. 5, pp. 1773-1796 , 2015.
- [7] Abson Karl, Palmer Ian, "Motion capture: capturing interaction between human and animal", Visual Computer, vol. 31 , No. 3, pp. 341-353,2015.
- [8] Pchelkin Stepan, Shiriaev Anton S., Freidovich Leonid B., Mettin Uwe, Gusev Sergei V., Kwon Woong, Paramonov Leonid, "A dynamic human motion: coordination analysis", Biological Cybernetics, vol. 109 , No. 4, pp. 657-680,2015.
- [9] Ylli K., Hoffmann D., Willmann A., Becker P., Folkmer B., Manoli Y., "Energy harvesting from human motion: exploiting swing and shock excitations", Smart Materials and Structures, vol. 24 , No. 4, pp. 657-680,2015, 24.
- [10] Hou Junhui, Chau Lap-Pui, Magnenat-Thalmann Nadia, He Ying, "Compressing 3-D Human Motions via Keyframe-Based Geometry Videos", IEEE Transactions On Circuits And Systems For Video Technology, vol. 25 , No. 4, pp. 51-62,2015.
- [11] Lan Rongyi, Sun Huaijiang, "Automated human motion segmentation via motion regularities", Visual Computer, vol. 31 , No. 4, pp. 35-53,2015.
- [12] Zhang Qiang, Zhang Shulu, Zhou Dongsheng, Keyframe "Extraction from Human Motion Capture Data Based on a Multiple Population Genetic Algorithm", Symmetry-basel, vol. 6 , No. 4, pp.926-937, 2014.
- [13] Royo Sanchez, Ana Cristina, Aguilar Martin, Juan Jose, Santolaria Mazo Jorge, "Development of a New Calibration Procedure and Its Experimental Validation Applied to a Human Motion Capture System", Journal of Biomechanical Engineering-transactions of the Asme, vol. 41 , No. 12, pp. 657-680, 2014.
- [14] Jiang Yinlai, Hayashi Isao, Wang Shuoyu, "Knowledge Acquisition Method Based on Singular Value Decomposition for Human Motion Analysis", IEEE Transactions on Knowledge and Data Engineering, vol. 12 , No. 26, pp. 3038-3050,2014.
- [15] Guan Qiuju, Li Caiyong, Guo Xuemei, Wang Guoli, "Compressive classification of human motion using pyroelectric infrared sensors", Pattern Recognition Letters, vol. 49 , No. 4, pp. 231-237,2014.
- [16] Moghaddass Ramin, Zuo Ming J., Liu Yu, Huang Hong-Zhong, "Predictive analytics using a nonhomogeneous semi-Markov model and inspection data", IIE Transactions, vol. 47 , No. 5, pp. 505-520, 2015.
- [17] Fort A., Mugnaini M., Vignoli V., "Hidden Markov Models approach used for life parameters estimations", Reliability Engineering & System Safety, vol. 41 , No. 4, pp. 85-91,2015, 136.
- [18] Qiao Yulong, Weng Lixiang, "Hidden Markov Model Based Dynamic Texture Classification", IEEE Signal Processing Letters, vol.22 , No. 4, pp. 509-512,2015.
- [19] Manandhar Achut, Torriero Peter A., Collins Leslie M., Morton Kenneth D., "Multiple-Instance Hidden Markov Model for GPR-Based Landmine Detection", IEEE Transactions on Geoscience and Remote Sensing, vol. 53 , No. 4, pp. 1737-1745,2015, 53(4):.
- [20] Lin Iris, Muser Erik, Munsell Michael, Benson Carmela, Menzin Joseph, "Economic impact of psychiatric relapse and recidivism among adults with schizophrenia recently released from incarceration: a Markov model analysis", Journal of Medical Economics, vol. 41 , No. 4, pp. 657-680,2015.
- [21] Durmaz Vedat, "Markov model-based polymer assembly from force field-parameterized building blocks", Journal of Computer-aided Molecular Design, vol. 29 , No. 3, pp. 657-680,2015.
- [22] Gong Zhaoning, Cui Tianxiang, Pu Ruiliang, Lin Chuan, Chen Yuzhu, "Dynamic simulation of vegetation abundance in a reservoir riparian zone using a sub-pixel Markov model", International Journal of Applied Earth Observation and Geoinformation, vol. 35 , No. 4, pp. 175-186,2015.
- [23] Gahrooei Mostafa Reisi, Work Daniel B., "Inferring Traffic Signal Phases From Turning Movement Counters Using Hidden Markov Models", IEEE Transactions on Intelligent Transportation Systems, vol. 41 , No. 4, pp. 657-680, 2015.