2016 3rd International Conference on Smart Materials and Nanotechnology in Engineering (SMNE 2016)
ISBN: 978-1-60595-338-0

A Classification Method of Human Motions Based on ApEn and RF

Jiayu Chen, Xinv Zhu
School of Reliability and Systems Engineering, Beihang University, Beijing, China

Dong Zhou, Chuan Lv
Science and Technology on Reliability and Environmental Engineering Laboratory & State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

ABSTRACT: According to the non-stationary, nonlinear and high dimensional characteristics of human motion caption data, a classification method for human motions based on approximate entropy (ApEn) and random forest (RF) is proposed in this paper. First, by preprocessing the motion capture data, key joints are selected to represent the motion instead of the all original joints. Second, the feature vector of each motion is extracted by calculating its ApEn. Third, the feature vectors are severed as the input vectors of the RF classifier for human motions recognition. Finally an experiment is conducted and the classification results achieve a high classification accuracy for 3 motion patterns (walk, run and jump), which confirms the highly accurate and effective performance of the proposed method.

1 INSTRUCTIONS

Generally, human motions consist of a hierarchy of themes, each of which is composed of a different set of spatiotemporal features (Etemad & Arya, 2010). In these themes, the primary ones are related to specific actions like walk, run and jump. Various variations and feature information on actions are contained and constantly present in the motion data. However, they are difficult for machine learning methods to extract and analyze due to their small spatiotemporal significance (Etemad & Arya, 2014), non-stationary, nonlinear and high dimensional characteristics. In this paper, a classification method for human motions based on approximate entropy (ApEn) and random forest (RF) is proposed to provide an effective and accurate classification technology for human motion patterns.

ApEn is proposed by Pincus (1991) and defined as the conditional probability of similar vectors' similarity when the vectors increase to m+1 dimensions from m dimensions. In this paper, ApEn is introduced to measure the complexity of time series, which has stronger anti-interfering and anti-noise ability. The ApEn of a motion contains the feature information and can express the characteristics of motions.

As a promising classifier, the RF classifier is a general term for ensemble methods using tree-type classifiers, which is induced by Breiman (2001). In the RF, Bagging is defined as a meta-algorithm to improve classification, and regression models according to stability and classification accuracy. Through using bagging, RF can build a large number of decision trees (Quinlan,1986, Yang, Park & Kim, 2000) out of a sub-dataset from a unique original training set and variance can be reduced, which helps to avoid over-fitting synchronously. This procedure selects cases randomly from original training data set and the bootstrap sets are used to construct each decision tree in the RF. Each tree classifier is named as a component predictor and the RF makes decisions by counting the votes of component predictors on each class. In this way, the winner class is selected in terms of the number of votes to it (Breiman, 2006).

Therefore, the paper is organized as follows. Section 2 introduces the details of the proposed method, consisting of selection of key joints, ApEn and RF. Experimental results are shown in section 3; Section 4 concludes this paper.

2 METHODOLOGY

2.1 Framework of the proposed methodology

As shown in the figure 1, the methodology has mainly three major steps: 1) Feature extracting process. In this step, capture data of human motions is analyzed and divided into different and standard motion clips, each of which contains a series of frames. Then, various joints in the frames are decomposed into several key joints. Lastly, by calculating the ApEn of motion clips, the feature is extracted to represent the motion clip. 2) RF training process. RF classifier is trained by the training samples in this step. 3) RF recognizing process. With the trained RF classifier and testing samples, the classification of human motions is conducted and the result is obtained.

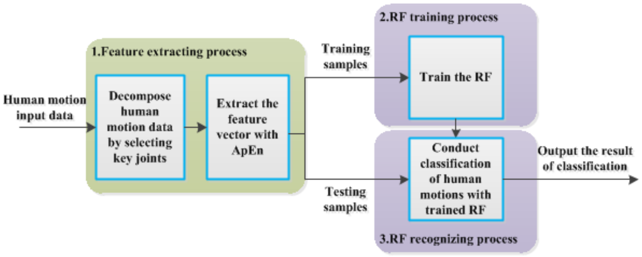

Figure 1. Framework of the proposed methodology.

2.2 *Selection of key joints*

There are two international standardization organizations in the field of modeling human. One is Moving Picture Expert Group and the other is Humanoid Animation Working Group (HAWG). In this paper, the capture data of human motions adopts Humanoid Animation Specification developed by HAWG (Gleieher, 2000), in which joints are to define joints' information of virtual human. These information contains data of position, name, translation, scaling, rotation, all of which are related to human motions. However, motion data of some joints changes little but increases the dimensions of the data, which makes it complicated to process the data and difficult to express the characteristics of a motion. A simple selection algorithm of key joints is applied to obtain the key joints and its principle is as follows:

1) Standardize a motion clip into a $n \times m$ matrix $M_{n,m}$, where n represents the number of frames, m represents the number of joints and $V_{n,m}$ represents the value in the row m and column n .

2) Set the number k of key joints and obtain former k maximum key joints by calculating $\max(V_i) - \min(V_i)$ (i=1, 2, 3,…, m).

3) Obtain the key joints matrix $M_{n,k}$ by using k maximum key joints replace all m joints.

2.3 *Approximate entropy*

A motion clip is consisted of a time series of n frames. For time series, ApEn is an effective tool to measure the complexity and extract the feature information of the motion clip. Algorithm for calculating the ApEn is described as follows:

1) For an N sample time series $\{u(i): 1 \le i \le N\}$, given m, form vector sequences $X_1{}^m$ through $X_{N-m+1}{}^m$ as

$$X_i{}^m = \{u(i), u(i+1), ..., u(i+m-1)\}, i = 1..., N-m+1 \quad (1)$$

Where m is the length of compared window.

2) For each $i \le N-m+1$, let $C_i{}^m(r)$ be $(N-m+1)^{-1}$ times the number of vectors $X_j{}^m$ within r of $X_i{}^m$. By defining

$$\phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i{}^m(r) \quad (2)$$

where $\ln$ is the natural logarithm.

3) For $m+1$, repeat above steps and compute $\phi^{m+1}(r)$. ApEn statistic is given by:

$$ApEn(m,r) = \lim_{N \to \infty} [\phi^m(r) - \phi^{m+1}(r)] \quad (3)$$

Through computing the ApEn, the feature is extracted to express the motion clip. Thus, ApEn can be applied to obtain the feature vector of a motion clip.

2.4 *Random forest*

Random forest is an ensemble classifier that is composed of many decision trees. It was first proposed to solve the classification problem (Breiman, 2001).

The RF classifier consists of a collection of tree structured classifiers $\{C(X, \theta_k), k = 1, ...\}$, in which θ_k is defined as the independent identically distributed random vectors and each decision tree casts a unit vote for the most popular class at input X.

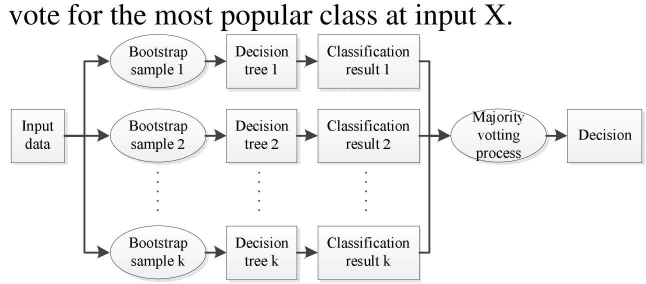

Figure 2. Framework of RF.

A general framework of training a RF classifier and classification process is shown as Figure 2 and described as follows:

1) By employing bootstrap sampling, k samples are selected from training set and the sample size of each selected sample is the same as the training sets.

2) k decision tree models are built for k samples and k classification results are obtained by these decision tree models.

3) According to k classification results, the final classification result is decided by voting on each record.

4) The differences among classification models are increased by building different training sets in a RF, therefore extrapolation forecasting ability of ensemble classification model is enhanced. After k times training, a classification model series $\{h_1(X), h_2(X), ..., h_k(X)\}$ is obtained, which is utilized to structure a multi-classification model system. The final classification result of the system is simple majority voting and the final classification decision is as Equation 4:

$$H(x) = \arg\max_Y \sum_{i=1}^{k} I(h_i(x) = Y) \quad (4)$$

Where $H(x)$ is the ensemble classification model, $h_i(x)$ is a single decision tree classification model, Y is the objective output, I is an indicative function, the equation 4 explains the final classification is decided by majority voting.

3 CASE STUDY

An experiment was conducted to illustrate the effectiveness of the proposed method. The capture data of human motion used in the experiment comes from Carnegie Mellon University Graphics Lab Motion Capture Database. The mocap lab in the basement of Wean contains 12 Vicon infrared MX-40 cameras, each of which is capable of recording 120 Hz with images of 4 megapixel resolutions. The cameras are placed around a rectangular area, of approximately 3m x 8m, in the center of the room. The database contains different representative motion patterns, such as walk, run, jump and so on. Walk, run and jump are the most popular and basic motion patterns, which cover more detailed patterns, such as jogging, forward jump, climb, swing. In this paper, three typical and basic motion patterns are selected: walk, run and jump.

The capture data is included in the asf/amc files, in which the data related to information on the actual markers is contained in the asf files, and information on the joint angles is reserved in the amc files. For classification of motion patterns, angle data in amc files is analyzed and used in the experiment.

3.1 Preprocessing motion capture data

Generally, the number of frames contained in a motion amc file varies from several hundred to several thousand according to the complexity of the motion. As basic motions, walk, run and jump can be expressed in a small amount of frames. For convenience of computation, the number of frames is set to 120 for motion walk and run. For motion jump, the number of frames is set to 240 because it spends more time to complete. Additionally, specific information on the joint angles in the amc files can be found in the table 1.

As shown in the table 1， each frame of motion data is consisting of 29 joints and 62 angles data. Thus, a standard walk, run and jump motion data can be expressed by a matrix $M^{W}_{120,62}$, $M^{R}_{120,62}$ and $M^{J}_{240,62}$.

Table 1. An example frame of joints' information in amc files.

Joints	Angle data
Root	7.5514 15.923 -40.3324 7.05373 -2.20224 -3.98785
Lowerback	2.60946 -0.102232 1.55785
Upperback	0.557657 -0.210319 3.00254
Thorax	0.897061 -0.104282 2.19971
Lowerneck	18.1347 -4.05186 -11.0833
Upperneck	18.0637 -5.11779 11.7943
Head	9.27447 -1.91673 5.26313
Rclavicle (Lclavicle)	-6.53505e-015 1.03368e-014 (-6.53505e-015 1.03368e-014)
Rhumerus (Lhumerus)	-37.2089 18.9892 -89.6591 (-46.3351 -13.9149 96.7412)
Rradius (Lradius)	39.7899 (30.1959)
Rwrist (Lwrist)	-20.0925 (7.25908)
Rhand (Lhand)	-30.1949 -24.0387 (-11.6462 -19.3885)
Rfingers (Lfingers)	7.12502 (7.12502)
Rthumb (Lthumb)	3.50572 -53.9896 (14.3979 10.0921)
Rfemur (Lfemur)	30.2197 -14.3203 4.47729 (-37.2769 11.458 -18.1673)
Rtibia (Ltibia)	27.5699 (68.9099)
Rfoot (Lfoot)	-17.5721 4.04349 (-11.1022 14.3613)
Rtoes (Ltoes)	-3.15233 (36.8211)

3.2 Results and discussions

360 feature vectors, consisting of 120 feature vectors of walk, run and jump respectively, were obtained through selection of key joints and computing ApEn. 90 feature vectors are randomly selected to compose the training set, which including 30 feature vectors from each class. Other feature vectors were set as the testing set.

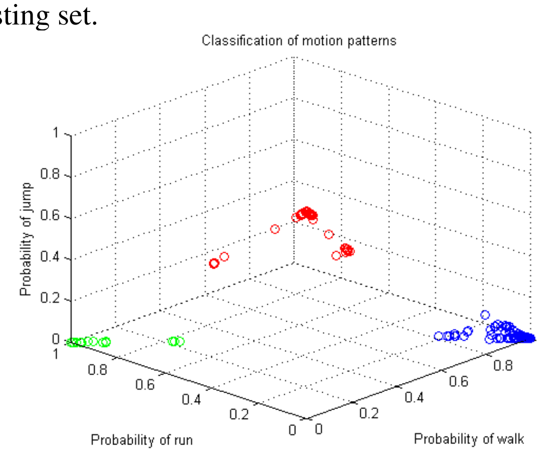

Figure 3. Classification results of motion patterns.

With 90 feature vectors in the training set, the RF classifier is trained. After 20 times classifications, the average results of 270 motions recognition are obtained and shown in the figure 3.

As shown in the figure 3，three typical motion patterns are classified and distributed in the three different and independent space, in which points with blue, green and red represents walk, run and jump respectively.

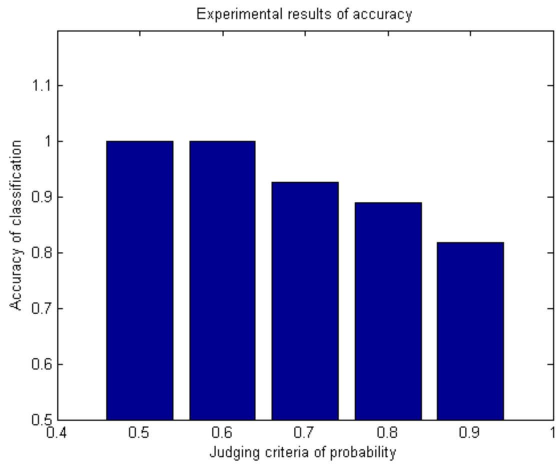

Figure 4. Experimental results of accuracy.

As shown in the figure 4, when the judging criteria of testing motions is that the probability obtained by the RF classifier is larger than 50%, 60%, 70%, 80% and 90%, the accuracy of classification is 100%, 100%, 92.6%, 88.9% and 81.9%.

As shown by the experimental results, it indicates that the proposed method has a significant performance in the classification of the three typical and basic motion patterns: walk, run and jump.

4 CONCLUSIONS

The paper proposes a classification method for human motion based on ApEn and RF. The capture data of human motion is non-stationary, nonlinear and high dimensional, which makes the classification of motion patterns difficult. Through selection of key joints, the original motion capture data is decomposed. Additionally, the ApEn of each motion is computed and the feature is extracted. Then, motion patterns can be recognized and distinguished by the RF classifier. Finally, an experiment is conducted. Using the proposed method, the complicated data is well processed and three similar motion patterns are classified with a high accuracy. The experimental results validate the feasibility and effectiveness of the proposed method.

A multitude of important efforts of future work is supposed to be dedicated to the classification of more typical motion patterns and even representation of human cognition according to recognition of human behaviors.

5 ACKNOWLEDGEMENT

Many thanks to graduate student innovation practice fund program of Beihang University for its found and support. Many thanks also to State Key Laboratory of Virtual Reality Technology and Systems, and Science and Technology on Reliability and Environmental Engineering Laboratory for providing the equipment and expertise to carry out the maintenance simulation experiment.

REFERENCES

[1] B. S. Yang, C. H. Park and H. J. Kim, An Efficient Method of Vibration Diagnostics For Rotating Machinery Using a Decision Tree, International Journal of Rotating Machinery, 6 (1) (2000) 19-27.

[2] Gleieher M. Animation from observation: motion capture and motion editing. Computer Graphics, 2000, 33(4): 51-54.

[3] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, (1986).

[4] L. Breiman, Random Forests, Machine Learning, 45 (1) (2001) 5-32.

[5] L. Breiman, Random Forest User Notes, Statistics Department, University of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/breiman/notes_on_random_forests_v2.pdf, (2006).

[6] Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America 1991; 88(6):2297–301.

[7] S. A. Etemad; A. Arya, Modeling and transformation of 3D human motion, in: Proceedings of the Fifth International Conference on Computer Graphics Theory and Applications, 2010, pp. 307–315.

[8] S. Ali Etemad, Ali Arya. Classification and translation of style and affect in human motion using RBF neural networks. Neurocomputing. 2014 129: 585-595.