doi:10.21311/002.31.4.02

# Action Recognition Based On Conceptors Of Skeleton Joint Trajectories

**Jiao Bao [1], Lishen Pei [2], Mao Ye[1], Xuezhuan Zhao[2]**

[1]School of Computer Science and Engineering,Center for Robotics, Key Laboratory for NeuroInformation of Ministry of Education,University of Electronic Science and Technology of China, Western High-Tech Industrial Zone, Chengdu, 611731, China;

[2]School of Computer Science, Zhengzhou University of Aeronautics, Zheng-Dong New Zone, Zhengzhou, 450046, China

## Abstract

With the tremendous popularity of the Kinect, recognizing human actions or gestures from skeletal data becomes more feasible. Skeletal data is a more exact data than RGB video while it eliminates the occlusions that caused by the limbs of the actor. Previous neural network based approaches recognize actions by learning spatial-temporal features. However, nobody can explain what are those features represent. Different from them, we propose a novel action recognition framework based on conceptors of skeleton joint trajectories. Conceptor is a mechanism of neurodynamical organization. We compute conceptor for the trajectory of each dimension of the skeleton joint, and use the singular value vector to represent the trajectory. Then, we encode singular value vectors as binary vectors by using a clustering method. At last, we use softmax regression to recognize the trajectory codes. This is a novel framework which recognizes actions using the conceptual level information. Extensive experiments on benchmark datasets confirm the efficiency of this framework.

**Keywords:** action recognition, conceptor, skeletal joint trajectories, softmax regression, recurrent neural network.

## 1. INTRODUCTION

Recognizing actions from human action videos or skeletal data gets much attention recent years. It enables wide applications such as intelligent surveillance, video understanding, human-computer interaction and smart home. Currently, most of the action recognition works (Dawn and Shaikh 2016; Pei et al., 2014; Pei et al., 2014; Pei et al., 2015; Pei et al., 2014) are focused on recognizing actions from RGB videos. Meanwhile, some research works recognize actions from multi-modal data, such as audio files, depth videos, RGB videos, skeleton, and so on. In most practical applications, multi-modal data is difficult and expensive to collect. With the popularity of the Kinect, collecting skeleton data of human actions becomes easier. To be robust to the confusions that caused by the limbs, clothes color or texture of the actor, recognizing actions from skeleton data (Jiang et al., 2014; Wang et al., 2016) is a wise choice.

With the development of deep learning, many neural network based approaches are proposed to recognize actions by learning spatio-temporal features. Those features are learned from the pixel data of the action videos. Although those features get excellent performance on action recognition, any literatures can not explain what are those features represent. This is different with human brain. The neural network is designed to processing data by imitating the human brain. While the human brain knows what it learns, the neural network has no idea on what it learns. Jrgen Schmidhuber says that

the human brain is a recurrent neural network (RNN). It can learn many behaviors that are not able to learn by traditional machine learning methods.

Another problem of those deep learning approaches is that they assume the input size and output size of the neural network are fixed. Affected by the action speed, behavior habit or action category, usually, the temporal lengths of the action sequences are not fixed. Despite images can be cropped or resized, action sequences are not amenable to a straightforward reduction for fixed size.

Some approaches are natural ways for sequence data processing. Such as Hidden Markov Model (HMM), RNN, etc. Those algorithms have no temporal size limitation. They are appropriate for action sequence processing. Combined with HMM, the Gaussian Mixture Model (GMM) (Murphy, 2012) and the Deep Belief Network (DBN) (Wu and Shao, 2014) are used to recognize actions from skeletal data. The two probability based approaches get good performance on action recognition. However, they can not remember the learned actions similar as the dynamical system RNN.

Based on RNN, a mechanism of neurodynamical organization is proposed, called Conceptor (Jaeger, 2014). It is possible to learn, store, abstract, generalize, de-noise and recognize dynamical patterns. We try to directly use the conceptor to regenerate the action skeletal sequence. It fails, because RNN have been shown to yield good performance for one dimensional sequences rather than high dimensional data (Palangi et al., 2013). When we use the conceptor to regenerate the trajectory of each dimension of the skeleton joints, it works. Based on this, we propose a novel framework to recognize actions from skeletal data.

At first, we compute a conceptor for the trajectory of each dimension of the skeleton joints. Then, the singular values of each conceptor are computed. For each dimension of the skeleton joints, all of the singular value vectors are clustered into several groups. Each singular value vector is encoded as a binary vector by k-means clustering algorithm.The skeletal actions are represented as vectors by combing those binary vectors. At last, we use softmax regression to recognize actions.The contributions of this method are multi-folds. First, at the conceptual level, we propose a novel framework to recognize actions.Second, different from other neural network based approaches which learns spatio-temporal features, we use the singular value vectors of the computed conceptors to represent the trajectories of the skeletal joints. Third, to reduce the dimension of the representation, we encode the singular value vectors as binary vectors using clustering algorithms. It is effective while using softmax regression to recognize actions.

The rest of this paper is organized as follows. After introducing the related works in Section 2, we give a detailed description of the proposed action recognition framework in Section 3. Then, the experiment validations are provided in Section 4. At last, Section 5 concludes this work.

## 2. RELATED WORKS

Skeletal data can be obtained by many approaches. Traditionally, the skeletal joint data are acquired by MoCap system (Muller and Roder, 2006). Currently, most of the skeletal data are acquired by the Kinect (Han et al., 2013). This device can extract human body joints in real time with reasonable accuracy. It has been found wide applications. With the development of the depth of computer vision research, some works (Toshev and Szegedy, 2014) focus on estimating skeleton from images. In this paper, we use the skeletal data that obtained by the Kinect to validate our approach.

Based on the action feature representation, the approaches that recognizing actions from skeletal data can be simply classified into two categories. One category is based on the frame-level features (Lv and Nevatia 2006; Muller Roder 2006; Wu and Shao 2014; Xia and Chen 2012; Yao et al., 2012). Most of such features are called pose features. The other category is based on the trajectory descriptors (Gowayyed et al., 2013; Ohn-Bar and Trivedi 2006; Qiao et al., 2015).The two category approaches have fast developments because of two reasons. In each frame, the skeleton joints have obvious geometric relationships that can be recorded machine learning techniques. For the whole action sequence, the skeletal joints are strictly corresponded over all the frames. So the trajectory of each joint can be easily and correctly achieved.

For the frame-level features based approaches, there are some classical features. As the harbinger of exploring skeletal joint data, (Muller and Roder, 2006) introduces the relational pose features for indexing and retrieving the motion capture data. Then, (Yao et al., 2012) modified the relational pose features to recognize actions. (Lv and Nevatia, 2016) designs a feature vector, the vector represents the pose of the combination of multiple joints or a single joint. (Wu and Shao, 2014) extracts the high level skeletal joints representation using the deep belief network. Take the quantized histogram of spherical coordinates of joint locations as frame-level features, (Xia and Chen, 2012) models the action sequence using HMM.

For the trajectory based methods, most approaches model the trajectory holistically. (Gowayyed et al., 2013) records the joint orientation displacements over the whole trajectory using a histogram. (Ohn-Bar and Trivedi, 2013) uses the pairwise affinities trajectories of joint angles to model actions. (Qiao et al., 2015) recognizes skeleton-based actions by learning discriminative trajectory detector sets. Mingxi Zhang (2016) proposed a moving body recognition model based on mean shift algorithm optimized by gradient features and conducted simulation experiment.

In this paper, we propose a novel action recognition approach based on the trajectory of each dimension of the skeletal joints. Different with the previous methods, it is a concept-level cognition system. Based on the conceptor mechanism, the actions can also be regenerated. Using a dynamic system with memory function to recognize action is a new attempt for action recognition.

## 3. METHODOLOGY

In this paper, we use the recurrent neural network that which is controlled by Conceptors (Jaeger, 2014) as a basic block to build an action recognition framework. At first, we give some preliminaries of Conceptors. Then, we provide a detail description of our action recognition framework. In addition, we split the whole framework into three procedures, and each procedure is demonstrated respectively.

### 3.1 Conceptor

Based on recurrent neural network, Herbert Jaeger uses neural filters that which are called Conceptors to characterize dynamical neural activation patterns. This approach has an excellent performance for one dimensional sequences. The architecture of standard recurrent neural network is shown in Fig. 1. In this figure, $p$ denotes the input dynamic pattern, $W_{in}$ and $W_{out}$ are the input weight matrix and the output weight matrix, $x$ denotes the activation states, $y$ is the output. The middle layer of the network is a reservoir. Its neurons are connected by random synaptic links. $W$ is the weight matrix of the reservoir.

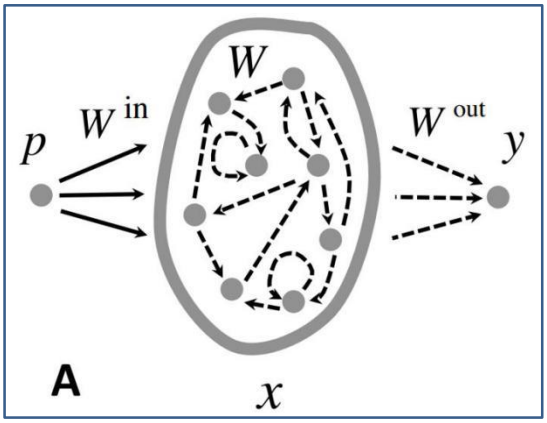


**Figure 1.** The architecture of standard recurrent neural network. The arrows present synaptic links.

Driving the reservoir using a dynamic patterns $p(n)$, the update function is shown as follows.

$$x(n+1) = \tanh(Wx(n) + W_{in}p(n+1) + b), \quad (1)$$

$$y(n) = W_{out}x(n). \quad (2)$$

Convenient for introducing our action recognition framework, we introduce the **Definition 1** and **the Proposition 1** of (Jaeger, 2014).

**Definition 1** Let $R = E[xx']$ be an $N \times N$ correlation matrix and $\alpha \in (0,\infty)$. The conceptor matrix $C = C(R,\alpha)$ associated with $R$ and $\alpha$ is

$$C(R,\alpha) = \arg\min_C E[\|x - Cx\|^2] + \alpha^{-2}\|C\|_{fro}^2. \quad (3)$$

**Proposition 1** Let $R = E[xx']$ be a correlation matrix and $\alpha \in (0,\infty)$. Then, $C = C(R,\alpha)$ can be directly computed from $R$ and $\alpha$ by

$$C(R,\alpha) = R(R + \alpha^{-2}I)^{-1} = (R + \alpha^{-2}I)R. \quad (4)$$

### 3.2 Network architecture

The architecture of our action recognition framework is shown in Fig. 2. Coarsely, the framework can be separated into two parts. The action representation procedure and the action classification procedure. In this section, we give a general description of our framework. With the general impression, the detailed introductions and analyses of the two procedures are shown in the following sections.

In Fig. 2, $(p_1, p_2, \ldots, p_n)$ represent the skeletal joint trajectories. The input weight $W_{in}$ and its corresponding bias are fixed with random values $W^*$. For each trajectory pattern, we use it to drive the reservoir. Update the reservoir states $x$ using Function 1. After a washout time, the reservoir states are collected in $x$. According the Definition 1 and Proposition 1, the conceptor $C$ are computed. Then, the singular values of the conceptor are computed. With a descending order, the singular values are represented as a

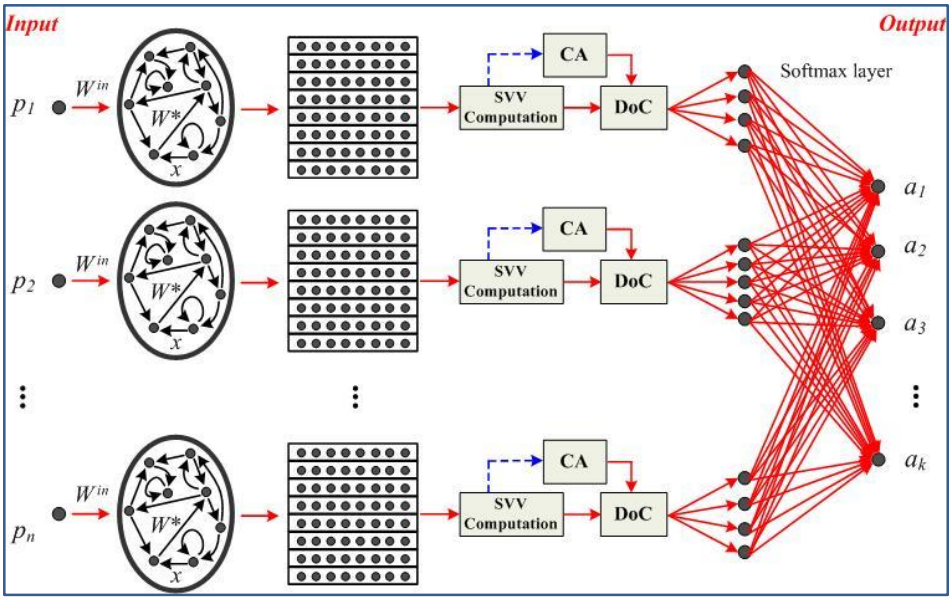


**Figure 2.** Action recognition framework based on Conceptors.

singular value vector. Using the clustering algorithms, the singular value vector are encoded as a binary vector. Combined with other trajectory coders, the action sequence are represented as a binary vector. At last, we use a modified softmax regression algorithm to recognize actions.

### 3.3 Representation of skeleton trajectories

Limited by the conclusion (Palangi et al., 2014) that echo state networks have shown to yield good performance on one dimensional sequences rather than high dimensional data. For the trajectory of a skeletal joint $P_i$, in the screen coordinates, it is denoted as $(p_{2i-1}, p_{2i})$. Based on the theory of conceptor, we compute the conceptor of the two dimensional dynamic pattern. However, many conceptors fail to regenerate their corresponding two dimensional patterns. For the trajectory of the skeletal joint, its temporal length, movement velocity and range have great arbitrariness because of the personal habits. Most of the current literatures are difficult to regenerate the action skeleton joint trajectories.

Considering the excellent ability of echo state network on one dimensional data processing, according to the dimension of the trajectory, each skeletal joint trajectory $P_i$ can separated into two trajectories $p_{2i-1}$ and $p_{2i}$. Then, we compute the conceptor of each trajectory. The conceptor contains enough information to regenerate (Jaeger, 2014) the trajectory of the action skeleton joint.

Using the singular value decomposition algorithm, we compute the singular values for each conceptor. With a descending order, the singular values are represented as a Singular Value Vector(SVV). In Fig. 3, the corresponding singular value vectors of those trajectories are depicted in the left columns of the six figures (a)-(e). In this paper, the coordinates of the skeletal joints are normalized to $(-1, 1)$. So, the trajectories of each dimension of the skeletal joints are one dimensional numerical sequences that varying over time in the range of $(-1, 1)$.

After collecting the singular value vectors of each trajectory dimension, We cluster them by using the $k$ -means algorithms. Each collection is clustered into $k$ groups. To encode a

singular value vector for the same trajectory dimension, we calculate the metric distances between the centers of the $k$ groups and the SVV. The distance vector $d^j = (d_1^j, d_2^j, ..., d_k^j)$ is used to represent the singular value vector of the corresponding trajectory that which is indexed by $j$. The whole skeletal action is encoded as $D = (d^1, d^2, ..., d^n)$ by combing those representations of the action joint trajectories. For simplicity, we denote $D = (d^1, d^2, ..., d^n)$ as $D = (d_1, d_2, ..., d_{k*n})$.

### 3.4 Action recognition

The human actions are dynamic patterns with uncertainty temporal length. As described above, based on the Conceptor techniques, the skeletal actions are encoded as feature vectors with definite dimension. For each skeletal action $S_i$ of the action training set, it is represented as feature vectors $D_i$. Then, we use a multi-class classifier softmax regression to recognize actions.

For each test action representation $D_i$, we use $p(y = j|D_i;\theta)$ to represent the estimated probability for each action class category $j$. Assume the action class number of one action data set is $m$, the hypothesis function $h_\theta(D_i)$ of softmax regression is as follows,

$$h_\theta(D_i) = \begin{bmatrix} p(y_i = 1|D_i;\theta) \\ p(y_i = 2|D_i;\theta) \\ \vdots \\ p(y_i = m|D_i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{m} \exp(\theta_j^T D_i)} \begin{bmatrix} \exp(\theta_1^T D_i) \\ \exp(\theta_2^T D_i) \\ \vdots \\ \exp(\theta_m^T D_i) \end{bmatrix}, \tag{5}$$

where $\theta_1, \theta_2, ..., \theta_m \in \mathbf{R}^{k*n+1}$ are the parameters of the softmax regression model, and $\sum_{j=1}^{m} \exp(\theta_j^T D_i)$ is the partition function. For simplicity, we use $\boldsymbol{\theta}$ to denote all of the model parameters for softmax regression. So, $\boldsymbol{\theta}$ is a parameter matrix with the dimension of $m \times (k * n + 1)$, and it can be denoted as follows,

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_m^T \end{bmatrix}. \tag{6}$$

All of the model parameters are learned by minimizing the cost function of the softmax regression algorithm. The cost function is shown as follows,

$$J(\theta) = -\frac{1}{n} \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} 1\{y_i = j\} \log \frac{\exp(\theta_j^T D_i)}{\sum_{l=1}^{m} \exp(\theta_l^T D_i)} \right], \tag{7}$$

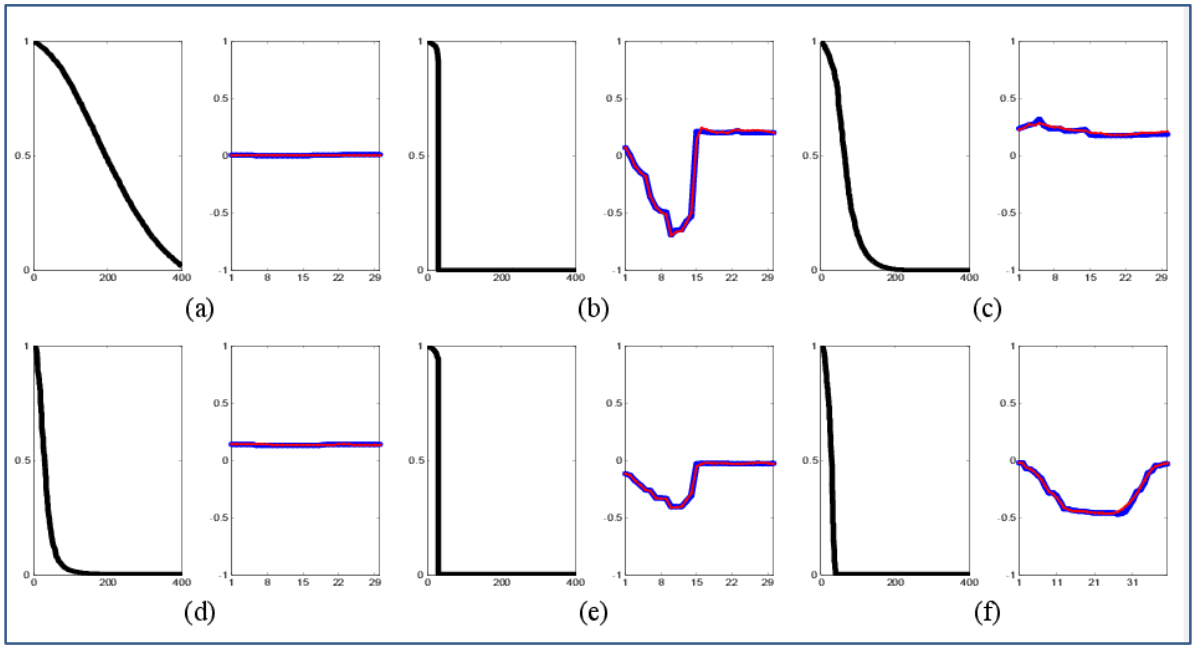


**Figure 3.** Some samples of the singular value vectors and their corresponding trajectories. For the right columns of figures (a)-(e), the blue lines are the trajectories of one dimension of skeletal joints, and the red lines are the regenerated trajectories that produced by Conceptors. The curves of the left columns of figures (a)-(e) are the computed singular value vectors of the right trajectories.

where $1\{.\}$ is an indicator function. With enough training action samples, we use L-BFGS algorithm to solve the optimization problem.

## 4. EXPERIMENTS

To validate the effectiveness of the proposed algorithm, we perform multiple experiments. First of all, we performed a regeneration experiment to demonstrate that, the conceptor contains enough information to regenerate the trajectory of each action skeleton joint. Then, we performed the action recognition experiment on two benchmark skeleton action datasets, ChaLearn Italian Gesture dataset and MSR Action3D dataset, and the proposed algorithm shows good experiment results.

### 4.1 Effectiveness Of the SVV representation

As demonstrated in (Jaeger, 2014), with the participation of conceptors, a recurrent neural network can regenerate different model patterns that which has previously been driven with. Using the algorithms that provided by (Jaeger, 2014), we regenerate the trajectories of each dimension of the action skeleton joints. In Fig. 3, some trajectories of one dimension of the skeletal joints (blue) and the regenerated trajectories (red) are depicted in the right columns of the six figures (a)-(e). We can find that the trajectories are correctly regenerated. It demonstrates that the Conceptor has enough information to regenerate the trajectory.

In Fig. 3, the trajectories of (b) and (e) have similar movement trends, the computed singular value vectors are similar with each other. (a) is a example of a stationary trajectory and (c) is a trajectory with slight movement, the singular value vectors are different. (a) and (d) are almost stationary trajectories of skeletal joints. However, the coordinate values of the two trajectories are different. The computed singular value vectors are different with each other. From those figures, we can find that, to a certain

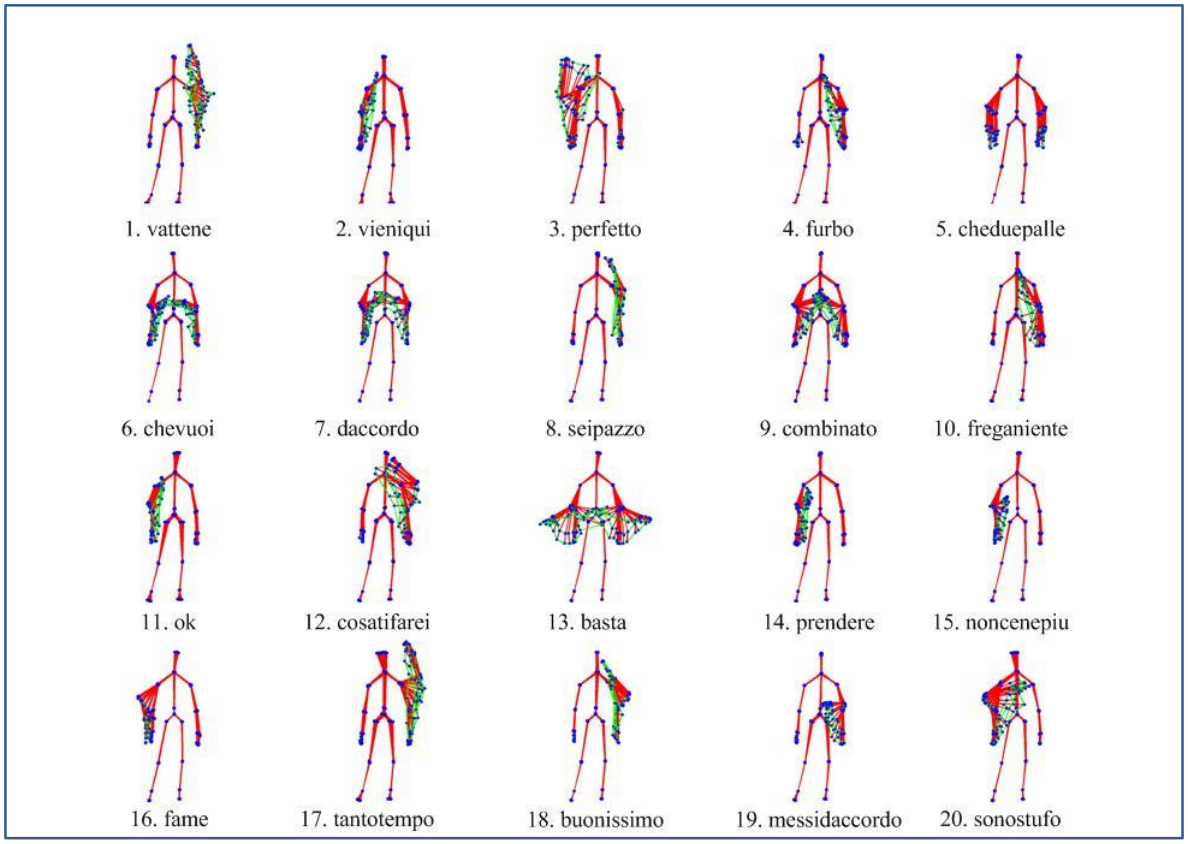


**Figure 4.** The 20 actions of the Chalearn Italian Gesture Dataset.

extent, those singular value vectors depict the trajectories of each dimension of the skeletal joints. For similar trajectories, the computed singular value vectors are similar with each other. Instead, the Singular Value Vectors have a great difference**.**

### 4.2 Action recognition experiment on ChaLearn Italian Gesture Dataset

The ChaLearn Italian Gesture Dataset (Escalera et al., 2013) is organized for the multi-modal gesture recognition challenge 2013. It is a multi-modal data set recorded by the Kinect camera. This dataset includes RGB video streams, depth images, user masks, skeletal data and audio data. In this paper, we only focus on the skeletal data. Each skeleton includes 20 joints. We perform experiments on the pixel positions of the skeletal joints. This dataset contains 20 Italian cultural/anthropological signs (Fig. 4). The names of those gestures are as follows, (1) vattene, (2) vieniqui, (3) perfetto, (4) furbo, (5) cheduepalle, (6) chevuoi, (7) daccordo, (8) seipazzo, (9) combinato, (10) freganiente, (11) ok, (12) cosatifarei, (13) basta, (14) prendere, (15) noncenepiu, (16) fame, (17) tantotempo, (18) buonissimo, (19) messidaccordo, (20) sonostufo.

To compare with other skeletal data based approaches, similar as (Wu and Shao, 2014), we perform experiments on a subset of this dataset. This set includes 393 sections. Each section contains between 8 and 20 gesture instances. In all, there are 7754 gestures for the experiments. Following the experiment setup of other approaches (Wu and Shao, 2014), we use the gestures of 350 sections to train the parameters, and the rest 43 sections are used for testing. With the above experiment setup, we performed the human action recognition experiment on the ChaLearn Italian Gesture dataset. While training the parameters, we use some copies of the training action sections with a little change of the action skeleton joint trajectories.

**Tabla 1** Average accuracy comparison on ChaLearn Italian Gesture Dataset.

| **Method** | **Accuracy** |
|---|---|
| GMM+HMM [Murphy, 2012] | 0.408 |
| EigenJionts [Yang and Tian, 2012] | 0.593 |
| NN+DTW [Wu et al., 2013] | 0.599 |
| DBN+HMM [Wu and Shao, 2014] | 0.628 |
| Our Method | 0.651 |

**Tabla 2** Average accuracy comparison on MSR Action3D Dataset.

| **Method** | **Accuracy** |
|---|---|
| Sequence of Most Information Joints [Ofli et al., 2014] | 0.29 |
| Recurrent Neural Network [Martens and Sutskever, 2011] | 0.425 |
| Dynamic Temporal Warping [Muller and Roder, 2006] | 0.54 |
| Hidden Markov Model [Lv and Nevatia, 2006] | 0.63 |
| Multiple Instance Learning [Ellis et al., 2013] | 0.657 |
| GMM+HMM [Murphy, 2012] | 0.704 |
| EigenJoints +DBNN [Yang and Tian,2012] | 0.72 |
| Structured Streaming Skeletons [Zhao at el., 2013] | 0.817 |
| DBN + HMM [Wu and Shao, 2014] | 0.82 |
| Our Method | 0.834 |

The action recognition result is shown in Table 1, and we compared the proposed action recognition algorithm with the state-of-the-art approaches that which are based on action skeleton information. The compared algorithms include the deep learning methods and the traditional machine learning approaches. Such as the deep learning algorithm(Wu and Shao, 2014) which uses deep belief neural network to process action skeleton information, and as the traditional machine learning algorithms (Murphy 2012; Wu et al., 2013; Yang and Tian 2012) which uses hidden markov model, naive-bayes-nearest-neighbor, or dynamic time warping to recognize actions. The proposed algorithm is different with them because that it import conceptors to represent action skeleton joint trajectories. Above all, the usage of conceptor improved the action recognition performance. From Table 1, it can be seen that the proposed algorithm get a better recognition performance.

### 4.3 Action recognition experiment on MSR Action3D Dataset

The MSR Action3D Dataset (Wang et al., 2012) is an action dataset that recorded by a depth sensor. The depth sensor is similar as the Kinect device. There are 20 action categories performed by 10 subjects in this dataset. Each action is performed 2 or 3 times by each subject. In this experiment, we only focus on the skeletal data of this dataset. There are 557 skeletal action sequences. Each skeleton has 20 joint positions. The joint position contains four real numbers: $u,v,d,c.$ $(u,v)$ are the screen coordinates, $d$ is the depth value, and $c$ is the confidence score. In this paper, we perform experiments on the $(u,v,d)$ coordinates of the skeletal joints. The normalization of the coordinates of the skeletal joints is the same as that which is introduced by (Wu and Shao, 2014).

This is a challenging dataset while some of the action skeleton information is very noisy. In this dataset, we use the cross-subject experiment setup to perform the action recognition experiments. The skeleton action of five subjects are used to train the parameters, and the actions of the other five subjects are used for testing. The average action recognition result is shown in Table 2, and we compared our results with some of

the state-of-the-art action recognition approaches. From this table, we can find that, the proposed algorithm get a better performance.

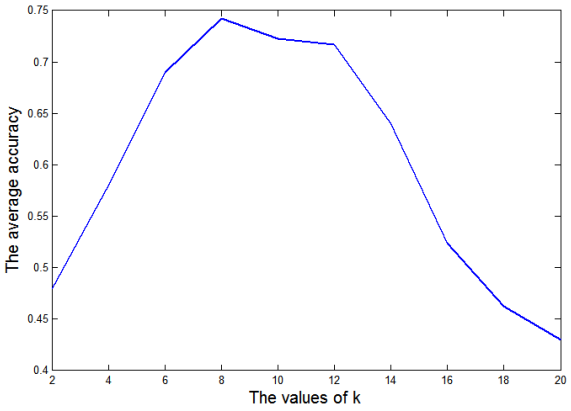


**Figure 5.** The average recognition accuracy of our method with different values of k which are used to generate binary vector for softmax.

### 4.4 Parameric analysis

In our proposed algorithm, the singular value vectors (SVVs) of the computed conceptor for the trajectory of of each dimension of the skeleton joint are encoded as a binary vector by using the clustering algorithms. When we generating these binary codes (vectors) for the SVVs, the variable k of the k-means clustering algorithm determines the dimension of the code. The larger k is, the more categories are classified. That is to say, it determines the dimension of input vector of the softmax regression. k becomes an indicator of the SVVs and the results of the final classification.

In order to discover the usefulness of the input binary code, our algorithm is tested with different values of k of the k-means clustering algorithm changing from 2-20 with 2 as the interval. The result is shown in Fig. 5. In Fig. 5, we compute the average accuracy of the action recognition based on two bench mark action datasets. As can be seen, too small or too large value of k may both lead to poor performance. Thus, a proper choice of k is important for good performance of our proposed algorithm. Experiments found that the value of k between 6-12 is a good choice.

## 5. DISCUSSION AND CONCLUSION

By importing conceptor, we proposed a novel action recognition approach based on RNN. As some of the current action recognition algorithms, we use the action skeleton information to process the actions. Compared with the RGB action videos or other formats of human actions, action skeleton is more precise, and it is very easy to avoid the occlusion problem that caused by the clothes, human body or the other things. However, limited by the action skeleton capture device, some captured skeleton actions are very noisy. So, it is still a challenging problem to recognize actions based on the skeleton information.

Currently, most of the action recognition algorithms through extracting spatial temporal features or learning the high level codes of actions to achieve recognize. The proposed algorithm is different with them. This algorithm computes a conceptor for each dimension of the action skeleton joint trajectory. The conceptors are not only a representation of the action skeleton joint trajectories. The conceptors contain enough information to regenerate the joint trajectories.It is also a mechanism of neurodynamical organization. The importation of the conceptor makes it is different with other feature

extracting or learning algorithms. The experiment results on two bench mark action datasets confirm the effectiveness of the proposed algorithm.

## 6. ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (61375038) and Applied Basic Research Programs of Sichuan Science and Technology Department (2016JY0088).

## 7. REFERENCES

Dawn D.D., Shaikh S.H. (2016). A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector, The Visual Computer, 32(3), 1-18, DOI: 0.1007/s00371-015-1066-2.

Ellis C., Masood S.Z., Tappen M.F., Jr., J.J.L., Sukthankar R. (2013). Exploring the trade-off between accuracy and observational latency in action recognition, International Journal of Computer Vision, 101(3), 420-436, DOI: 10.1007/s11263-012-0550-7.

Escalera S., Gonzlez J., Bar X., Reyes M., Lopes O., Guyon I., Athistos V., Escalante H.J. (2013). Multimodal gesture recognition challenge 2013: Dataset and results, ACM ChaLearn Multi-Modal Gesture Recognition Grand Challenge and Workshop, 445-452, DOI:10.1145/2522848.2532595.

Gowayyed M.A., Torki M., Hussein M.E., ElSaban M. (2013). Histogram of oriented displacements (HOD): Describing trajectories of human joints for action recognition, IEEE International Joint Conference on Articial Intelligence, 1351-1357.

Han J., Shao L., Xu D., Shotton J. (2013). Enhanced computer vision with microsoft kinect sensor: A review, IEEE Trans. On Cybernetics, 43(5), 1318-1334, DOI: 10.1109/TCYB.2013.2265378.

Jaeger H. (2014). Controlling recurrent neural networks by conceptors, Jacobs University Technical Report Nr 31.

Jiang X., Zhong F., Peng Q., Qin X. (2014). Online robust action recognition based on a hierarchical model, The Visual Computer, 30(9), 1021-1033, DOI: 10.1007/s00371-014-0923-8 .

Lv F., Nevatia R. (2006). Recognition and segmentation of 3d human action using HMM and multi-class adaboost, European Conf. On Computer Vision, 3954, 359-372, DOI: 10.1007/11744085_28.

Martens J., Sutskever I. (2011). Learning recurrent neural networks with hessian-free optimization, IEEE Conf. on Machine Learning, 1033-1040.

Mingxi Zhang (2016), Optimized moving body behavior recognition model based on multi-texture gradient feature, Revista Tecnica De La Facultad De Ingenieria Universidad Del Zulia, 39 (5), 299-305, DOI: 10.21311/001.39.5.39.

Muller M., Roder T. (2006). Motion templates for automatic classification and retrieval of motion capture data, ACM SIGGRAPH/Eurographics Symposium on Computer animation, 137-146, DOI: 10.2312/SCA/SCA06/137-146.

Murphy K.P. (2012). Machine learning: a probabilistic perspective, The MIT Press, 58(8), 27-71, DOI: 10.1038/217994a0.

Ofli F., Chaudhry R., Kurillo G., Vidal R., Bajcsy R. (2014). Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition, Journal of Visual Communication and Image Representation, 25(1), 24-38, DOI: 10.1109/CVPRW.2012.6239231.

Ohn-Bar E., Trivedi M.M. (2013). Joint angles similiarities and HOG2 for action recognition, In Proc. Workshops of IEEE Conf. on Computer Vision and Pattern Recognition, 465-470, DOI: 10.1109/CVPRW.2013.76.

Palangi H., Deng L., Ward R.K. (2013). Learning input and recurrent weight matrices in echo state networks, IEEE Conf. on Neural Information Processing Systems.

Pei L.S., Ye M., Xu P., Li T. (2014). Fast multi-class action recognition by querying inverted index tables, Multi-media Tools and Applications, 74(23), 1081-10822. DOI: 10.1007/s11042-014-2207-8.

Pei L.S., Ye M., Xu P., Zhao X.Z., Guo G. (2014). One example based action detection in hough space, Multimedia Tools and Applications, 72(2), 1751-1772, DOI: 10.1007/s11042-013-1478-9.

Pei L.S., Ye M., Zhao X.Z., Dou Y.M., Bao J. (2014). Action recognition by learning temporal slowness invariant features, The Visual Computer, 1-10, DOI: 10.1007/s00371-015-1090-2.

Pei L.S., Ye M., Zhao X.Z., Xiang T., Li T. (2014). Learning spatio-temporal features for action recognition from the side of the video, Signal, Image and Video Processing, 1-8, DOI: 10.1007/s11760-014-0726-4

Qiao R., Liu L., Shen C., Hengel A.V.D. (2015). Learning discriminative trajectory let detector sets for accurate skeleton-based action recognition, IEEE Conf. on Computer Vision and Pattern Recognition, 1-10.

Toshev A., Szegedy C. (2014). Deeppose: Human pose estimation via deep neural networks, IEEE Conf. on Computer Vision and Pattern Recognition, 1653-1660, DOI: 10.1109/CVPR.2014.214.

Wang C., Wang Y., Yuille A.L. (2016). Mining 3D key-pose-motifs for action recognition, IEEE Conf. on Computer Vision and Pattern Recognition, 2639-2647.

Wang J., Liu Z., Wu Y., Yuan J. (2012). Mining actionlet ensemble for action recognition with depth cameras, IEEE Conf. on Computer Vision and Pattern Recognition, 36(50), 1290-1297, DOI: 10.1109/CVPR.2012.6247813.

Wu D., Shao L. (2014). Leveraging hierarchical parametric networks for skeletal joints based action segmentation and recognition, IEEE Conf. on Computer Vision and Pattern Recognition, 724-731, DOI: 10.1109/CVPR.2014.98.

Wu J., Cheng J., Zhao C., Lu H. (2013). Fusing multi-modal features for gesture recognition, In proceeding of the 15th ACM on International Conference on Multi-modal Interaction, 453-460, DOI: 10.1145/2522848.2532589.

Xia L., Chen C.C., Aggarwal J. (2012). View invariant human action recognition using histograms of 3D joints, In Proc. Workshops of IEEE Conf. on Computer Vision and Pattern Recognition, 20-27, DOI: 10.1109/CVPRW.2012.6239233.

Yang X., Tian Y. (2012). Eigenjoints-based action recognition using naive-bayes-nearest-neighbor, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 38(3C), 14-19, DOI: 10.1109/CVPRW.2012.6239232.

Yao A., Gall J., Gool L.V. (2012). Coupled action recognition and pose estimation from multiple views, IEEE International Journal of Computer Vision, 100(1), 16-37, DOI: 10.1007/s11263-012-0532-9.

Zhao X., Li X., Pang C., Zhu X., Sheng Q.Z. (2013). Online human gesture recognition from motion data streams, ACM international conference on Multimedia, 23-32, DOI: 10.1145/2502081.2502103.