Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **3**, May **2016**

A Human Behavior Recognition Method Based on Latent Semantic Analysis

He-Jin Yuan, Chun-Hong Duo, Wei-Hua Niu

Department of Computer
North China Electric Power University
071003 Baoding, China
yhj_1977@163.com

Received June, 2015; revised December, 2015

Abstract. *In this paper, the image sequence of human behavior is regarded as a "document", and the key postures are regarded as "words" in the "document". Then, the latent semantic analysis method, which has obtained very good performance in natural language processing, is utilized to analyze and model the relationships among the image sequences and the key postures. In our method, we firstly calculate the mesh features of each image in human behavior sequence. Then the mesh features are vector quantized through a rival penalized competitive neural network and the behaviors described by time-sequential images are converted into symbolic sequences. Through above processing, the image sequences of human behavior can be looked as "documents" composed by the key postures, which is represented by the competitive neurons. Then, we use latent semantic analysis method to model the relationships among the behaviors and the key posture. The redundancy and noise can be effectively removed through singular value decomposition. Finally, the observed behavior is classified by the maximum posteriori probability criteria. The experiments on Weizmann and KTH datasets demonstrate that our method is effective.*

Keywords: Human behavior recognition; Latent semantic analysis; Competitive neural network

1. **Introduction.** Human behavior recognition based on vision is analysis and recognition of human behavior in video sequence, and it has wide application in visual surveillance, intelligent sensing interface, content-based video retrieval and other fields. Human behavior recognition essentially is a time-varying signal recognition problem. The existing recognition methods can be divided into two types: the methods based on template matching and the methods based on probability network. The template matching method converts image sequence into one or a set of templates, then matches the behavior to the known templates. Bobick and Davis[1] put forward a human action recognition algorithm with two temporal templates, named motion energy image and motion history image. Wang[2] gave an action recognition algorithm with average motion energy and mean motion shape templates. Weinlan[3] presented an action recognition approach using exemplar-based embedding. The template matching methods can be simply implemented, but they are sensitive to the time interval of the behavior. And they needs time warping before matching. The methods based on probability network define each static human posture as a state, and connect these states through the network, then use probability to describe the transitions between states. The commonly used probability networks are hidden Markov model, dynamic Bayesian network and conditional random

fields, in which the hidden Markov model is a stochastic model which is used mostly. In Yamato's method[4], the human images are divided into equal meshes and the pixels of each mesh are used as the feature vector for behavior recognition. Ryoo[5] also uses it for gesture recognition. Although this method is capable of modeling the slight variation in the temporal and spatial scales of human behavior, it needs large number of labeled samples to learning the parameters. And the learning process is also very complex.

The key of human behavior recognition is how to efficient represent the behaviors and effectively measure the similarity between different behaviors. Human behavior in video sequences is a dynamic procedure. It is not only related to each frame's body posture, but also related to the order and duration of these postures. Even the same kind of behavior, different individuals will be different due to the variation of human body height, size and so on. This paper is inspired by latent semantic analysis in natural language processing. The image sequence of human behavior is regarded as a document, and the key human posture is regarded as words in the document. The potential relationship between image sequences and key posture is analyzed and modeled, and a human action recognition method based on latent semantic analysis is proposed.

The rest of this paper is organized as follows: A brief introduction about latent semantic analysis is presented in section 2. The details about human behavior recognition based on LSA, such as the feature representation of human behavior, the quantization coding of human behavior feature vectors, human behavior modeling based on latent semantic analysis and human behavior classification based on Bayesian maximum posteriori criterion are given in Section 3. In section 4, we evaluate our method with well known Weizmann and KTH datasets before concluding in section 5.

2. **Latent Semantic Analysis.** Latent semantic analysis (LSA) is a statistical model. It is widely used for knowledge acquisition, induction and expression in natural language processing. Latent semantic analysis looks each document as a point of lexical space coordinates. And the documents are not randomly distributed in this space. Their distribution has some semantic structure. Similarly, each word can also be seen as a point which is based on a document of spatial coordinate system. Through latent semantic analysis, we can model the dual probability relationship between words and documents in the semantic space. LSA can establish a latent semantic space through singular value decomposition. In this space, words and documents are projected on different dimensions and each dimension represents a latent conception. And then, we can extract semantic relationships between the words and obtain their semantic structure.

The key idea of latent semantic analysis is mapping the words and documents into a low-dimensional vector space, i.e. latent semantic space. The specific details are as follows:

For a given matrix $X_{m\times n}$, suppose its rank is r .Then X can be decomposed into two orthogonal matrixes and a diagonal matrix, i.e. $X = TSD^T$ .Among them,$T_{m\times r} = (t_1, t_2, ..., t_r)$ is a orthogonal matrix and $t_1, t_2, ..., t_r$ is the left eigenvectors of X. And they are also eigenvectors of matrix XX^T actually. $S_{r\times r} = diag(\sigma_1, \sigma_2, ..., \sigma_r)$ is a diagonal matrix and $\sigma_1, \sigma_2, ..., \sigma_r$ are the singular values of X . They are also the square roots of eignvalues of matrix XX^T or X^TX . And $\sigma_1, \sigma_2, ..., \sigma_r$ satisfies the relation: $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$. $D_{n\times r} = (d_1, d_2, ..., d_r)$ is another orthogonal matrix and $d_1, d_2, ..., d_r$ are right eigenvectors of X . And they are eigenvectors of matrix X^TX actually.Therefore, the matrix X can be expressed as:

$$X = \sigma_1 t_1 d_1^T + \sigma_2 t_2 d_2^T + ... + \sigma_r t_r d_r^T \tag{1}$$

The largest k singular values are reserved in latent semantic analysis. Through this processing ,the main framework of the semantic space is retained and the noises included in the samples represented by matrix X are thus eliminated. The specific details are as follows: let $S_k = diag(\sigma_1, \sigma_2, ..., \sigma_k)$,$T_k = (t_1, t_2, ..., t_k)$, $D_k = (d_1, d_2, ..., d_k)$, then $\hat{X}^k = T_k S_k D_k^T$. And $\hat{X}^k$ is the best approximation of matrix X in the sense of least square of error. $T_k = (t_1, t_2, ..., t_k)$ is K base vectors in latent semantic space, and each base vector represents a dimension of latent semantic space. The base vectors also can be regarded as latent concepts or hidden attributes. Generally, it is difficult to find the true meaning of these latent concepts. They only have the significance on the statistical probability.

Assume $doc_i^* = (doc_{i,1}^*, doc_{i,2}^*, ..., doc_{i,k}^*)^T$ and $doc_j^* = (doc_{j,1}^*, doc_{j,2}^*, ..., doc_{j,k}^*)^T$ is the low-dimensional representation of two documents in latent semantic space, the similarity between these two documents can calculated by correlation coefficient or cosine between doc_i^* anddoc_j^*. The correlation coefficient formulas is as follows:

$$sim(doc_i^*, doc_j^*) = \sum_{h=1}^{k} doc_{i,h}^* doc_{j,h}^* \tag{2}$$

And the formula for cosine correlation is as follows:

$$sim(doc_i^*, doc_j^*) = \frac{\sum_{h=1}^{k} doc_{i,h}^* doc_{j,h}^*}{\sqrt{\sum_{h=1}^{k} \left(doc_{i,h}^*\right)^2} \sqrt{\sum_{h=1}^{k} \left(doc_{j,h}^*\right)^2}} \tag{3}$$

In the latent semantic space created by LSA, the semantics of the words or documents are described through the combination of latent conceptions. And this combination is not symbols, but linear algebra. Among them, the hidden conception corresponding to the large singular value represents more common property and less individual characteristics. In order to reflect the different dimension difference, we can make the important dimensions play more roles in the comparison of document vectors. So, we need to weight the dimensions when measuring the similarity between documents. Here, we use formula (4) to define the weight of different dimensions:

$$doc_i' = (\sigma_1 d_{1,i}, \sigma_2 d_{2,i}, ..., \sigma_k d_{k,i})^T \tag{4}$$

In above formula, the singular values are defined as the weights of the corresponding dimensions. And the weighted document vectors can be expressed as:$doc^* = doc^T T_k$.

3. **Human Behavior Recognition Based on LSA.** The overall procedure of human behavior recognition algorithm based on latent semantic analysis is shown in Figure 1 .It mainly includes human behavior feature extraction, clustering, coding, human behavior modeling based on latent semantic analysis and behavior classification based on the maximum posteriori probability criterion.

3.1. **The Feature Representation of Human Behavior.** In case of the target segmentation is accurate, human behavior can be distinguished by the shape feature of human body regions in each frame. So, silhouette based features are widely used in behavior recognition since they can be easily and robustly extracted from videos. Fig.2 shows the

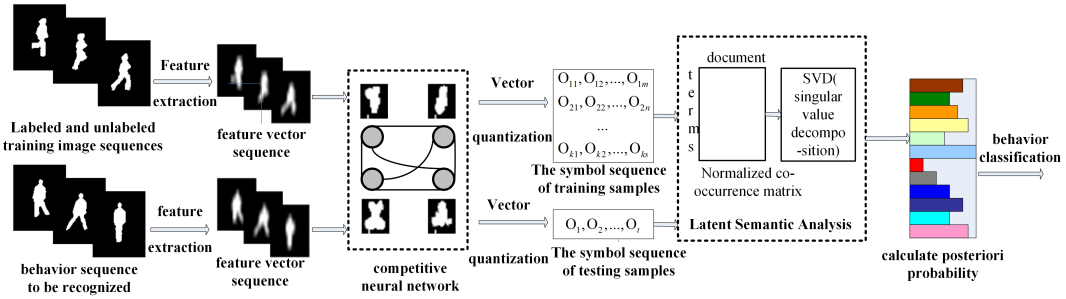


FIGURE 1. The overall procedure of human behavior recognition algorithm based on LSA

(a) The human body silhouette sequence of running behavior in Weizmann dataset

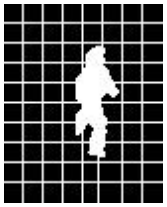
(b) The human body silhouette sequence of running behavior in Weizmann dataset

(c) The key body postures of bending and running

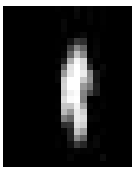
FIGURE 2. The shape feature examples of human behavior

human body regions in each frame of bend and run series in Weizmann[7] dataset. Obviously, the human body shapes in bend and run actions are very different. As shown in Fig.2(c), the key pose silhouettes of human body in bend and run actions are also very different. Furthermore, the number of similar silhouette and their occurrence order reflect the temporal and dynamic information of the behaviors, such as speed and duration time.

Silhouette based shape descriptors commonly include invariant moment, zernike moment and wavelet moment. However, these moments are computational and sensitive to noise image segmentation, which may be unavoidable because of the complex background in natural scene. So, in this paper, we use mesh features proposed in [4] as the human action representation for its lower computation and robustness for disjoint and inaccurate silhouette. Specifically, the image after segmentation is normalized, and then divided into grids. The proportion of the human body's pixels contained in each grid represents the



(a) Mesh feature



(b) The image representation of the mesh features of human behavior

FIGURE 3. The statistical image feature of human behaviors

body feature of this frame. According to the knowledge of human anatomy, the image should be divided into head, trunk and limbs and other non molecular region, and the pixel proportion is extracted from each sub region. For simplicity, in this paper we directly divide the image into equal grids with size of $M \times N$, as shown in Figure 3(a), and calculate the proportion of white pixels in each grid as the representation of human behavior. As shown in Figure 3 (b), the brighter the area is, the more pixels it contains.

3.2. **Mesh Feature Vector Quantization.** In this paper, we want to use latent semantic analysis to model the relationships among the behaviors and human body postures. So, here we need to change the image sequences of human behaviors into symbolic sequences through vector quantization .In addition, quantization coding for human mesh features can also shield body differences and noises caused by segmentation.

During quantization, the granular selection is a very important problem. If the granular is too small, the actions with same category label may be dissimilar; while if granular is too large, the detailed difference between actions may be shield. Since competitive neural network has the advantages of robustness and the on-line learning ability, here we use it for vector quantization. However, how to select an appropriate number of output neurons and avoid the influence of weight initialization are two difficult problems in competitive neural network learning. For these problems, Xu had put forward an effective method named rival penalized competitive learning (RPCL) algorithm [8].Its basic idea is that for each input, not only the weights of the winner unit are modified to adapt to the input, but also the weights of its rival are delearned by a smaller learning rate.

The basic steps of RPCL algorithm are as follows:

1) Selecting a relatively larger number K (the number of competitive neurons), and initializing the weights of the competitive neurons;

2) Choosing a sample x from the training set randomly, and calculating the following formula for $i = 1, 2, \cdots, K$:

$$u_i = \begin{cases} 1 & \text{if} \quad i = c \quad \text{such} \quad \text{that} \quad \gamma_c \|\mathbf{X} - \mathbf{W}_c\|^2 = \min_j \gamma_j \|\mathbf{X} - \mathbf{W}_j\|^2 \\ 1 & \text{if} \quad i = r \quad \text{such} \quad \text{that} \quad \gamma_r \|\mathbf{X} - \mathbf{W}_r\|^2 = \min_{j, j \neq c} \gamma_j \|\mathbf{X} - \mathbf{W}_j\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Here, $\gamma_j = {}^{n_j}/_{\sum_{i=1}^{K} n_i}$ and n_j is the cumulative number of $u_i = 1$. The import of γ_j overcomes the "dead node" problem, by which the influence of the neuron weight initialization is eliminated

3) Adjusting the weight of the competitive neuron according to the following formula:

FIGURE 4. The image representation of the competitive neuron weights

$$\Delta w_i = \begin{cases} a_{\mathrm{c}}(x - w_i) & \text{if} \quad u_{\mathrm{i}} = 1 \\ -a_{\mathrm{r}}(x - w_i) & \text{if} \quad u_{\mathrm{i}} = -1 \\ 0 \quad \text{otherwise} \end{cases} \tag{6}$$

$$w_i = w_{i-1} + \Delta w_i \tag{7}$$

Here,$0 \leq a_r, a_c \leq 1$ are the learning rates for the winner and rival unit respectively and it should be hold that $a_r << a_c$ in practice.

4) Running step 2) and 3) repeatedly until the maximal iteration or the weights of the neurons don't change obviously.

When using RPCL algorithm to train the neural network, a relative larger (the number of neurons) is given at the beginning. With the development of the training, the redundant neurons will be repelled far from the training data. After the algorithm ends, re-labeling the training data and deleting the neurons, which only have only very small amount of training data corresponding to them. Then the remainder neurons will be the final result.

For the human behaviors in Weizmann[7] dataset, the image representation of partial competitive neurons are shown in Fig.4,obviously they are comprehensive of key postures in different individual behavior. For each mesh feature vector extracted from the action images, calculating its distances to all the neurons, and encoding this feature vector as the number of the neuron, which is nearest to the current mesh feature vector. Through this processing, the time-sequential images of human action can be converted into symbol sequences.

3.3. **Human Behavior Modeling Based on Latent Semantic Analysis.** In this paper, we use latent semantic analysis to model and recognize human behaviors. The specific details are as follows: the key postures of human body represented by competitive neurons are regarded as "words" in LSA and the symbolic sequence of human behaviors are regarded as "documents". Suppose the ith behavior after mesh feature quantization is $d_i = (t_1, t_2, ..., t_{li})$. Here,$t_1, t_2, ..., t_{li}$ are the codes of each frame in the video sequence and li is the frame number of the video sequence. For each behavior sequence, counting the occurrence number of each competitive neurons to obtain the "lexical representation" of the behavior. That is $d_i^* = (s_1, s_2, ..., s_K)$ .And $s_1, s_2, ..., s_K$ respectively represents the occurrence number of competitive neurons in the ith behavior sequence. Through above procedure, the human behavior training set composed by video sequences can be converted into a matrix $X_{m\times K}$. Here, m is the number of video sequences. Decomposing the matrix $X_{m\times K}$ by LSA, we can obtain the left and right singular vector: $T_k = (t_1, t_2, ..., t_k)$, $D_k = (t_1, t_2, ..., t_k)$ and the singular value $\sigma_1, \sigma_2, ..., \sigma_r$. Then the human behaviors can be represented as $\hat{X}^k = T_k S_k D_k^T$.

3.4. **Human Behavior Classification Based on Bayesian Maximum Posteriori Criterion.** Suppose the human action sequences included in training set are $P_i(i = 1, 2, ..., n)$, and the class label of P_i is $C(P_i)$. Suppose the prior probability of actions with class label i is $\pi_i(i = 1, 2, \cdots, K)$, which can be simply determined by their percent in the training set. Then, the probability of the observed action T to class i is $P(i|\mathrm{T}) = \pi_i \times$

$P(T|i)/P(T)$ according to Bayesian theorem. Here, $P(T|i)$ is the conditional probability. Since its accurate calculation is difficult, here we simply suppose it is proportional to the matching degree between T and the action which is most similar to T in the training set with class label i , i.e.

$$P(T|i) \propto sim(P_s, T) \tag{8}$$

$$P_s = \arg\max_{C(P_j)=i}\{Sim(T, P_j)\} \tag{9}$$

Here, we use formula (3) to calculate the similarity between behaviors. And P_s is the behavior which is most similar to T in the training set with class label i after LSA. Then, the posteriori probability can be calculated as:

$$P(i|T) = \frac{\pi_i \times P(T|i)}{\sum_{j=1}^{K} \pi_j \times P(T|j)} \tag{10}$$

So, the final classification result can be determined through maximal posteriori probability criteria as follows:

$$C(T) = \arg \max_{i=1,2,\ldots,K} P(i|T) \tag{11}$$

4. **Experimental results and analysis.**

4.1. **Dataset.** In order to evaluate the proposed algorithm, we use two publicly available datasets Weizmann[7] and KTH[8]. Weizmann dataset is recently widely used in human action recognition algorithm, and contains 10 kinds of human behaviors, such as bend, jack, pjump, jump, run, side, skip, walk, wave1 and wave2. Each behavior has 9 performers, and run, skip and walk in Lena have two sequence behaviors as illustrated in Fig.5.

KTH dataset contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors s1, outdoors with scale variation s2, outdoors with different clothes s3 and indoors s4 as illustrated in Fig.6.

In our experiments, all recognition rates were computed with the leave-one out cross validation.

4.2. **Results and Analysis.** The experiments results on the Weizmann dataset are summarized as follows: the recognition accuracy of our method is 92.50%, the confusion matrix is shown in Table 1. As can be seen, it is prone to confusion among jump, skip, side and walk, wave1 and pjump since these actions are more similar than others. In comparison, the recognition rate of exemplar-based embedding method reported in [3] is 97.7% for 50 exemplars. The work of Ali et al. in [10] used a motion representation based on chaotic invariants and reported 92.6%, while Wang and Suter reported recognition rate of 97.78% with an approach that uses kernel-PCA for dimensional reduction and factorial conditional random fields to model motion dynamics in [2]. Table 2 summarizes action classification accuracies using different schemes on Weizmann dataset.

The experiments results on KTH dataset are summarized as follows: the recognition accuracy of our method is 86.5%, the confusion matrix is shown in Table 3. In comparison, the recognition rate of 3-D Harris+3-DHOG reported in [14] is 88.3% for 100 times per

FIGURE 5. Human action types in Weizmann video library

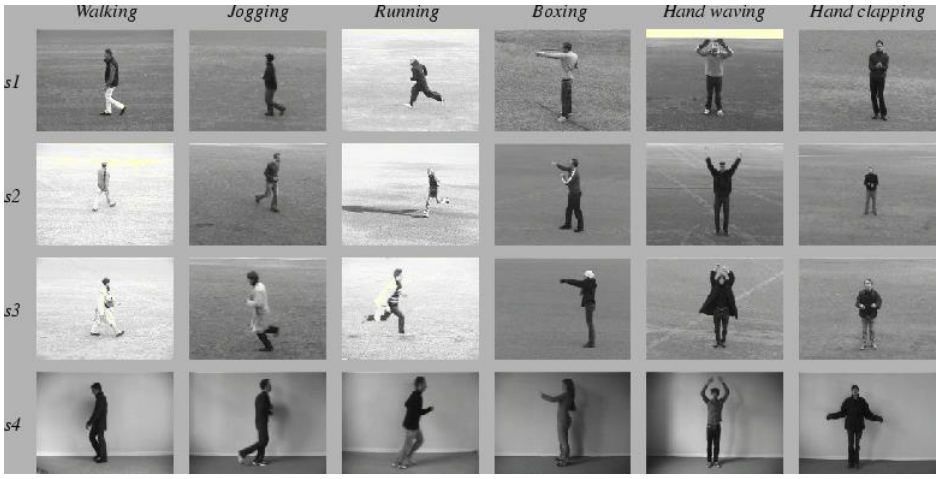


FIGURE 6. Human action types in KTH video library

action. The work of Dollar P et al. in [15] used the method of Cuboid Detector and Cuboid Descriptor reports 80%, while Willems G reported recognition rate of 84.26% in [16], and Klaser A reported recognition rate of 91.8% in [17]. Table 4 summarizes action classification accuracies using different schemes on KTH dataset.

The accuracy of our method is very close to those of state-of-art approaches. However, comparing to the existed approaches, our method is much easier to be implemented and has less parameters needed to be adjusted.

In order to test the efficiency of the algorithm, we use VC++6.0 to implement the program in the computer of Pentium Dual Core@2.70GHz and 2 GB memory. The Weizmann dataset has 5679 frames, the number of iterations is 500, the learning time of the competitive neural network is 180.3s, and the average classification time of each sample to be tested is 0.093s. In ref[14], the time of Cuboid Detector+Cuboid Descriptor is 0.164s, the time of 3-D Harris+HOG/HOF is 0.096s, and the time of 3-D Harris+3-D HOG is 0.088s. So, the average classification time of our method is very close to them.

5. **Conclusion.** This paper presented a new human behavior recognition method based on latent semantic analysis. The method first calculates the mesh features of each image in human behavior sequence. Then, the mesh features are vector quantized and the behaviors described by time-sequential images are converted into symbolic sequences. Finally, the latent semantic analysis method is utilized to model the relationships among

Table 1. The confusion matrix of recognition results on Weizmann dataset

	bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2
bend	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
jack	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
jump	0.00	0.00	0.89	0.00	0.00	0.00	0.11	0.00	0.00	0.00
pjump	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
run	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
side	0.00	0.00	0.00	0.00	0.00	0.89	0.11	0.00	0.00	0.00
skip	0.00	0.00	0.10	0.00	0.00	0.00	0.90	0.00	0.00	0.00
walk	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.90	0.00	0.00
wave1	0.11	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.78	0.00
wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.89

Table 2. the recognition rate statistics of Weizmann dataset

Recognition methods	rate (%)
Method based on the embedded samples [3]	97.70
Method based on kernel PCA feature reduction and conditional random field modeling [2]	97.78
Method based on chaotic invariant representation [10]	92.60
Method based on shape hierarchy model [11]	72.80
Method based on space-time shape model [12]	97.50
Method based on Bootstrap characteristics behavior [13]	98.30
Method in this paper	92.50

Table 3. The confusion matrix of recognition results on KTH dataset

	boxing	handclapping	handwaving	joggingh	running	walking
boxing	0.91	0.00	0.00	0.04	0.03	0.02
bandclapping	0.00	0.93	0.06	0.01	0.00	0.00
handwaving	0.00	0.05	0.91	0.01	0.01	0.02
jogging	0.02	0.00	0.01	0.82	0.09	0.06
running	0.02	0.00	0.00	0.10	0.80	0.08
walking	0.01	0.00	0.01	0.09	0.07	0.82

Table 4. the recognition rate statistics of KTH dataset

Recognition methods	rate (%)
3-D Harris+3-DHOG[14]	88.3
Cuboid Detector+Cuboid Descriptor[15]	80
Hessian+ESURF[16]	84.26
3-D Harris+HOG/HOF[17]	91.8
Method in this paper	86.5

the behaviors and the key postures and the observed behavior is classified by the maximum posteriori probability criteria. Compared to the existing methods, it has the following characteristics: 1)the human behavior sequence is looked as a "document", and the key posture is regarded as "words" in the "document";2)the potential relationships among image sequences and key postures are analyzed and modeled by latent semantic analysis method; 3) through clustering the mesh feature with competitive neural network , it can not only extract out the key-pose silhouette feature of different actions, but also can shield the influence caused by different actors' body shape difference or segmentation noise.

However, there are still many problems needed further research. First of all, although the results on Weizmann and KTH dataset are encouraging, evaluations on larger and realistic database need to be investigated in order to be more conclusive. In addition, the variation of camera orientation and zoom must be considered in the future.

Acknowledgment. The work reported in this paper was supported by "the Fundamental Research Funds for the Central Universities (2014MS129, 2014MS133)". We also acknowledge the anonymous reviewers for comments that lead to clarification of the paper.

REFERENCES

[1] A. F. Bobick, J. W. Davis, The Recognition of Human Movement Using Temporal Templates, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 257–267, 2001.

[2] L. Wang, D. Sute, Recognizing Human Activities From Silhouettes: Motion Subspace and Factorial Discriminative Graphical Mode, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–7, 2000.

[3] D. Weinland, E. Boyer, Action Recognition Using Exemplar-based Embedding, IEEE International Conference on Computer Vision and Pattern Recognition, pp.1–7, 2008.

[4] J. Yamato, J. Ohya, K. Ishii, Recognizing Human Action in Time-sequential Images Using Hidden Markov Model, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 379–385, 1992.

[5] S. RyooM, J. K. Aggarwal, Recognition of Composite Human Activities Through Context-free Grammar Based Representation, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1709–1718, 2006.

[6] S. Deerwester, S. T. Dumais, G. W. Furnas, et alm, Indexing By Latent Semantic analysis, *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[7] http://www.wisdom.weizmann.ac.il/ vision/SpaceTimeActions.html

[8] http://www.nada.kth.se/cvap/actions/

[9] L. Xu, A. Krzyzak, E. Oja, Rival Penalized Competitive Learning for Clustering Analysis, RBF Net and Curve Detection, *IEEE Trans. on Neural Networks*, vol. 4, no. 4, p. 636– 649, 1993.

[10] S. Ali, A. Basharat, M. Shah, Chaotic Invariants for Human Action Recognition, *International Conference on Computer Vision*, pp. 1–8, 2007.

[11] J. C. Niebles, F. F. Li, A Hierarchical Model of Shape and Appearance for Human Action Classification, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[12] L. Gorelick, M. Blank, E. Shechtman, Actions as Space-time Shapes, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 29, pp. 2247–2253, 2007

[13] C. Liu, P.C. Yuen, Human Action Recognition Using Boosted Eigen Actions, Image and Vision Computing, no. 28, 825–835, 2010.

[14] B. W. Zhang, Research of Human Active Recognition Based on Video, Chongqing University, Chongqing, China , pp. 47–52, 2014.

[15] P. Dollar, V. Rabaud, G. Cottrell, et al, Behavior Recognition via Sparse Spatio-Temporal Features, *2nd Join IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 66–72, 2005.

[16] G. Willems, T. Tuytelaars, An Efficient Dense and Scale-Invariant Spatio-Temporal Interest Point Detectors, Springer Berlin Heidelberg, pp. 650–663, 2008.

[17] A. Klaser, M. Marszalek, C. Schmid et al, A Spatio-Temporal Descriptor Based on 3D-Gradients, *BMVC*, pp. 1–10, 2008.