Tensor Analysis and Multi-Scale Features Based Multi-View Human Action Recognition

Chengcheng Jia, Sujing Wang, Xiangli Xu, Chunguang Zhou, Libiao Zhang
College of Computer Science and Technology, Jilin University, Changchun 130012, P. R. China

***Abstract*—A method of multi-view human action recognition based on multi-scale features via tensor analysis is proposed. A series of silhouettes are transformed to a Serials-Frame image, from which the multi-scale features are extracted to construct the eigenSpace of a tensor, which named Serials-Frame Tensor(SF-Tensor). The SF-Tensor subspace analysis is applied to separate the variable views and people information to recognize different actions. Experiment results obtained show that the proposed method attains a good recognition rate and improves the efficiency significantly.**

Keywords- action recognition; multi-scale features; SF-tensor; multi-view

I. INTRODUCTION

Vision-based human action recognition is currently one of the most active research in the domain of computer vision due to its vast application, such as automated surveillance system, sign language interpretation, analysis of sports events and many other movement-based human-computer interaction(MB-HCI) applications. So far the vision-based pose recognition has been extensively studied, such as hand, arm, gait and full-body poses[1]. A pose is the unit of an action, e.g. waving a racket, while an action is an ordered series of several movements[2], each of which has some kinds of relevance. Recently there have been more action recognition methods than before using probability and statistics and artificial intelligence technology such as Dynamic Bayes Network(DBN)[3], Hidden Markov Models[4] and grammar analysis. Existing methods can also be categorized according to the single-view[5] vs. multi-view[6], and types of features extracted e.g., 2D silhouette[5] and 3D volumetric reconstruction[1],[6].

It is important to obtain multiple views for vision-based action recognition in many MB-HCI applications, because a series of silhouettes of the same action have great differences from variable views. The Multi-layer DBN model[3], star skeleton[4] and Energy-Action model[7], all of which had not been considering the multi-view factor which is carrying weight in the process of recognition. Recently the multiple views method has been used in action recognition[1], facial recognition[8], sign language recognition and so on. One challenge for multi-view methods is that precise external camera calibration is usually required for 3D reconstruction. In addition, matching the observed body shapes to templates possibly in different orientations remains a challenging research problem[1].

A number of multi-view action recognition algorithms have been developed based on silhouettes. For example, recently pose recognition[1] and similarly gait recognition[9] are achieved by tensor analysis with the silhouettes extracted from the 2D set of poses. There are many other feature extraction methods such as velocity[3], period of movement, human star skeleton[4] and Energy-Action model[7] etc.

A multi-view action recognition method is proposed in this paper based on multi-scale features and tensor analysis, by putting a Serials-Frame image into the eigenSpace of a tensor named Serials-Frame Tensor (SF-Tensor). A Serials-Frame image constituted by a series of silhouettes is proposed in order to represent an action composed of continuous relevant poses. In fact, multiple scales of motion details are contained in the human activities. The motion related to moving trace mainly reflect the situation between people and surroundings or interaction of multi-people, and this motion is analyzed on the big scale. The motion related to poses such as bending and waving hand often is applied on the medium scale. The motion related limbs mostly is used on small scale[3]. We proposed a recognition method based on the multi-scale features extracted from the three scales, then a Serials-Frame Tensor (SF-Tensor) is used to recognize a real action consisting of a series of relative poses. The multi-scale features under multiple views can be put into a SF-Tensor which is produced by the core tensor and an action basis matrix, a view basis matrix, a people basis matrix and a Serials-Frame basis matrix using Higher- Order SVD (HOSVD)[10]. Given a query input of one of the target actions, corresponding action and view coefficient vectors can be computed using the core tensor. By analyzing this SF-Tensor, an unknown action could be classified under multi-view conditions. In order to distinct the dangerous actions from the ordinary actions, the movements of upper-limbs have been considered emphatically. The rest of this paper is organized as follows: in Section II, the definitions related to tensor are described; in Section III, the multi-scale features extraction and Serials-Frame Tensor algorithm are introduced; in Section IV, the experimental results are provided and conclusions are drawn in Section V.

II. TENSOR FUNDAMENTALS

A tensor is a multidimensional array. More formally, an Nth-order tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system. In this paper, lowercase italic letters (a, b,...) denote scalars, bold lowercase letters (a, b,...) denote vectors, bold uppercase letters (A, B, ...) denote matrices, and calligraphic uppercase letters ($\mathcal{A}$, $\mathcal{B}$, ...) denote tensors. Following the formal definition is given below:

Definition 2.1: The order of a tensor $\mathcal{A} \in R^{I_1 \times I_2 \times \ldots \times I_N}$ is N . The mode-n vectors (fibers) of $\mathcal{A}$ are the I_n - dimensional

Corresponding author: Libiao Zhang
Email address: lbzhang@jlu.edu.cn.

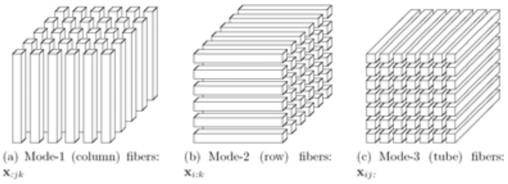


Figure 1. Fibers of $\mathcal{A}$ 3rd-Order Tensor

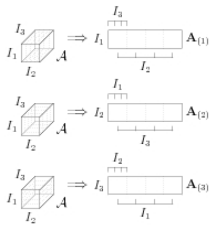


Figure 2. Flattening of the $(I_1 \times I_2 \times I_3)$ -tensor $\mathcal{A}$ to the $(I_1 \times I_2 I_3)$ - matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$ -matrix $\mathbf{A}_{(2)}$, and the $(I_3 \times I_1 I_2)$ -matrix $\mathbf{A}_{(3)}$ $(I_1 = I_2 = I_3 = 4)$

vectors obtained from A by fixing every index but index i_n

The mode-n vectors of $\mathcal{A}$ is given in Fig. 1 and the flattened matrix of A is given in Fig. 2.

Definition 2.2: The mode-n product of a tensor $\mathcal{A} \in R^{I_1 \times I_2 \times \ldots \times I_N}$ by a matrix $\mathbf{U} \in R^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an ($I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N$)-tensor of which the entries are given by

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} \overset{\text{def}}{=} \sum_{i_n} a_{i_1 i_2 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} u_{j_n i_n} . \quad (1)$$

This mode-n product of tensor and matrix can be expressed in terms of unfolding matrices for ease of usage.

$$(\mathcal{A} \times_n \mathbf{U})_{(n)} = \mathbf{U} \cdot \mathbf{A}_{(n)} \quad (2)$$

Given the tensor $\mathcal{A} \in R^{I_1 \times I_2 \times \ldots \times I_N}$ and the matrixes $\mathbf{U} \in R^{J_n \times I_n}, \mathbf{V} \in R^{J_m \times I_m}$, one has

$$(\mathcal{A} \times_n \mathbf{U}) \times_m \mathbf{V} = (\mathcal{A} \times_m \mathbf{V}) \times_n \mathbf{U} = \mathcal{A} \times_n \mathbf{U} \times_m \mathbf{V} \quad (3)$$

Similarly, N-mode SVD is a generalization of the SVD for higher order matrixes[11]. If $\mathcal{D}$ is an n-order tensor and $\mathcal{D} \in R^{I_1 \times I_2 \times \ldots \times I_N}$, the application of n-mode SVD orthogonalizs "n" associated vector spaces of $\mathcal{D}$ and decomposes the tensor as

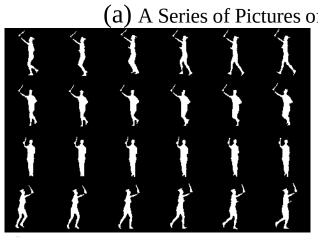
$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_n \mathbf{U}_n \ldots \times_N \mathbf{U}_N \quad (4)$$

where $\mathbf{U}_n$, $\forall n \in \{1, 2, \ldots, N\}$, is an orthonormal matrix and contains the ordered principal components for the n-mode. $\mathcal{Z}$ is called the core tensor. The decomposition algorithm is as follows:

1) For $n = 1, \ldots, N$, compute matrix $\mathbf{U}_n$ in (4) by computing the SVD of the flattened matrix $\mathbf{D}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD.

2) Solve for the core tensor as follows:

(a) A Series of Pictures of Multi-View



(b) A Series of Silhouettes of Multi-View

(c) Serial- Frame Image(SFI) of Multi-View

Figure 3. Multi-View Serials-Frame Image

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_n \mathbf{U}_n^T \ldots \times_N \mathbf{U}_N^T \quad (5)$$

III. PROPOSED APPROACHES

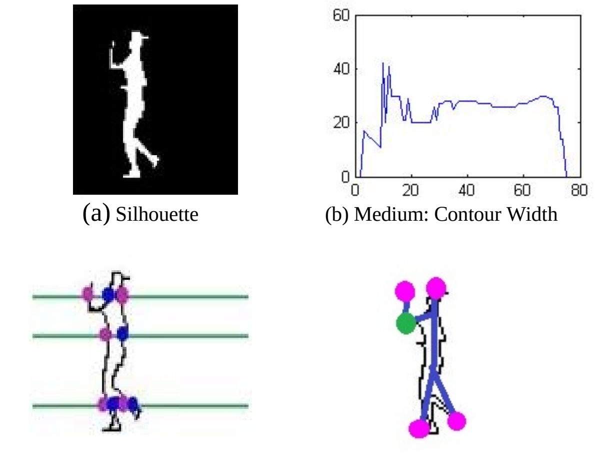
A. *Serials-Frame image(SFI) construction*

Distinction from extracting a silhouette as feature, a series of silhouettes were used to represent one action as a Serial-Frame image(SFI). See Fig. 3. Firstly a certain number(16) continuous silhouettes of one action were given into I_f as columns of the Serial-Frame matrix $\mathbf{A}$, then the series of silhouettes were resized to 120×160 to reduce dimension, thirdly every silhouette was cropped to a uniform size 80×80 as the dimensions of one-frame in the subspace I_{fea} when using silhouette features, see Fig. 4(a). In this way, a $I_f \times I_{fea}$ Serials-Frame matrix $\mathbf{A}$ is for every SFI, and the relevant subspace is represented as $R^{I_f \times I_{fea}}$.

B. *Multi-Scale features(MSF) from Serials-Frame image*

A number of pose recognition algorithms have been developed based on Silhouette Images(SI) attained as features. For example, in[12]pose recognition was achieved by comparing observed silhouette with the silhouettes projected from the 3D models of a set of poses. In[13]a recognition system was proposed based on the "envelope shape" representation of poses; experiments were conducted on simple actions. Recently, the Gaussian mixture models(GMM) have been used to represent pose silhouettes and the Kullback-Leibler divergence to compute distances between silhouettes. Bo Peng[1] used tensor analysis based on silhouettes to do pose recognition.

Using silhouette features for action recognition, there are 6400 dimensions in the space I_{fea} because the size of a silhouette is 80 × 80, on this condition the dimensions of the



(a) Silhouette (b) Medium: Contour Width

(c) Medium Intersectant points (d) Small: Key Points

Figure 4. Multi-Scale Features

eigenSpace $I_{(f)} \times I_{(fea)}$ of a tensor are too magnitude. Instead of taking a silhouette as feature, a method of extracting Multi-Scale features (MSF) from a Serials-Frame image as the dimension of I_{fea} is proposed. First of all the silhouettes were converted into contours for preprocessing. converted into contours for preprocessing

1) Big-Scale: On the big scale, the velocity of moving human was used as feature. Firstly the center of a contour was calculated, then the velocity could be attained by the time and distance of the variant centers in a series of continuous frames of a Serials-Frame image.

2) Medium-Scale: On the medium scale, the contour was used to attain the medium-scale feature, which was composed by three parts. Firstly, the width of each contour was extracted by raster-scanning, which reflected the scope of the movements of the upper-limbs and legs mainly. Then the number of the intersectant points was obtained by scanning the contour line by line, which reflected the osition of the upper-limb. Finally the period of the swing upper-limb was attained, which could be the measurement of moving frequency. The mediumscale feature consisting of the three parts has 161 dimensions.

Fig. 4(b) shows the contour width of a silhouette, it is obviously to see that the upper-limb has been moving in a larger scope compared to other parts on the body. Fig. 4(c) shows that the number of intersectant points on every line crossed with the contour, it is easily to see that the number of intersectant points of the upper-limb area is more than that of other areas.

3) Small-Scale: On the small scale, the positions of key points of human skeleton were used to express the action. The points are composed by 5 end points[14] consisting of head and four limbs and 1 corner point[15] of swing upper-limb from a human skeleton, so the small scale feature has 12 dimensions. Fig. 4(d) shows that the key points of human skeleton, and it is obviously to see that the limb might not be detected if there is obstruction to some extent. The multi-scale 1 + 161 + 12 = 174 dimensions vector is used as one-frame feature of a Serials-Frame image, noted as , and the eigenSpace of the tensor is $R^{I_f \times I_{fea}}$.

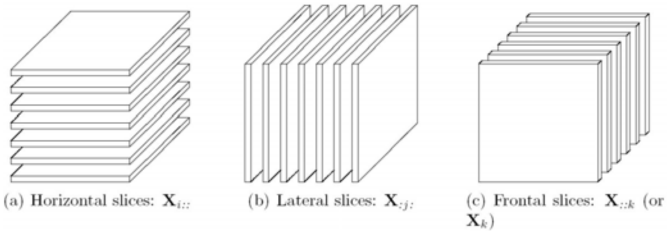


Figure 5. Hyper-slices of A 3rd-Order Tensor

C. Serials-Frame based tensor (SF-Tensor)

Tensor-Action is a multi-linear extension of a Serials-Frame action image. Given a Serials-Frame image including multiple factors such as action, view and people, a tensor $\mathcal{D} \in R^{I_a \times I_v \times I_p \times I_f \times I_{fea}}$ can be constructed, where I_a, I_v, I_p, I_f and I_{fea} denote the dimensions of action, views, people, serials of frames and one-frame feature in a Serials-Frame image respectively. HOSVD is applied to decompose the action, view and people information etc. as

$$D = Z \times_1 U_a \times_2 U_v \times_3 U_p \times_4 U_f \times_5 U_{fea} \tag{6}$$

where the core tensor $\mathcal{Z} \in R^{I'_a \times I'_v \times I'_p \times I'_f \times I'_{fea}}$ governs the interaction between the factors represented in the 5 mode matrices. The mode matrix $\mathbf{U}_a \in R^{I_a \times I'_a}$, $\mathbf{U}_v \in R^{I_v \times I'_v}$,and $\mathbf{U}_p \in R^{I_p \times I'_p}$ represent the parameters space of various action, views and people, respectively. The mode matrix $\mathbf{U}_f \in R^{I_f \times I'_f}$ and $\mathbf{U}_{fea} \in R^{I_{fea} \times I'_{fea}}$ constitute the eigenSpace of a Serials-Frame image. According to Equation (3), Equation (6) can be transformed follow Equation (7):

$$\mathcal{D} = (\mathcal{Z} \times_2 \mathbf{U}_v \times_3 \mathbf{U}_p \times_4 \mathbf{U}_f \times_5 \mathbf{U}_{fea}) \times_1 \mathbf{U}_a = \mathcal{B} \times_1 \mathbf{U}_a \tag{7}$$

D. Recognition Using SF-Tensor

Before discussing SF-Tensor algorithm, we introduce two formulated descriptions. As an extend of fibers, we can definite mode-n hyper-slices.

Definition 3.1: The mode-n hyper-slices of N-order $\mathcal{A} \in R^{I_1 \times I_2 \times \ldots \times I_N}$ could be decomposed as below:

$$\mathfrak{S}_{(I_n)}(\mathcal{A}) = \{\mathcal{A}_{:\ldots:1\ldots::}, \mathcal{A}_{:\ldots:2\ldots::}, \ldots, \mathcal{A}_{:\ldots:I_n\ldots::}\} \tag{8}$$

The mode-1 hyper-slices, mode-2 hyper-slices and mode-3 hyper-slices of a 3rd-order tensor $\mathcal{X}$ are denoted by $\mathbf{X}_{i::}, \mathbf{X}_{:j:}$ and $\mathbf{X}_{::k}$, respectively, see Fig. 5. The mode-n hyper-slices for mode-n product has the following properties.

Property 3.1: Given the tensor $\mathcal{A} \in R^{I_1 \times I_2 \times \ldots \times I_{n-1} \times I_n \times I_{n+1} \times \ldots \times I_N}$, $\mathcal{B} \in R^{I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$, the matrixes $\mathbf{U} \in R^{J_n \times I_n}$ and $\mathcal{B} = \mathcal{A} \times_n \mathbf{U}$.

TABLE I. MULTI-VIEW RECOGNITION RATE OF EACH METHOD

	MSF SF-Tensor	*SI SF-Tensor*	*MSF Clustering*	*SI Clustering*
view1 (0°)	65.00%	57.50%	34.38%	37.50%
view2(45°)	60.00%	45.00%	46.88%	40.63%
view3(90°)	55.00%	35.00%	37.50%	28.13%
view4(135°)	60.00%	60.00%	43.75%	31.25%
view5(180°)	55.00%	67.50%	40.63%	31.25%
view6(225°)	60.00%	55.00%	40.63%	28.13%
view7(270°)	77.50%	25.00%	40.63%	25.00%
view8(315°)	65.00%	47.50%	34.38%	28.13%

Applying hyper-slices operator along I_v, I_p on $\mathcal{D}$ and $\mathcal{B}$ in Equation (7), respectively, we can get

$$\mathfrak{S}_{(I_v)}(\mathfrak{S}_{(I_p)}(\mathcal{D})) = \{\mathcal{D}_{:vp::}\} \tag{9}$$

$$\mathfrak{S}_{(I_v)}(\mathfrak{S}_{(I_p)}(\mathcal{B})) = \{\mathcal{B}_{:vp::}\} \tag{10}$$

where $v = 1,\ldots,I_v; p = 1,\ldots,I_p$.

$\mathcal{B}_{:vp::} \in R^{I_a \times 1 \times 1 \times I_f \times I_{fea}}$ is the basis tensor for a particular viewpoint v and people p. According to Equation (7), we can obtain

$$\mathcal{D}_{:vp::} = \mathcal{B}_{:vp::} \times_1 \mathbf{U}_a \tag{11}$$

where $\mathcal{D}_{:vp::}, \mathcal{B}_{:vp::} \in R^{I_a \times 1 \times 1 \times I_f \times I_{fea}}$, $\mathbf{U}_a \in R^{I_a \times I_a}$. Give an unknown action image, a $1 \times 1 \times 1 \times I_f \times I_{fea}$ tensor $\mathcal{D}_{test}$ can be constructed. The action can be determined by following:

$$\arg\min_{a,v,p} \left\| \mathcal{D}_{test} - \mathcal{D}_{avp::} \right\| \tag{12}$$

IV. EXPERIMENTS

A. *The action dataset*

There are 4 different actions in the dataset, which are running with a tool lifted, walking, running while hitting with bare hand, walking while waving hand. Each of the actions consists of a series of 16 continuous frames in a video section, which is performed by 10 people in 8 different torso views of which are: $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ respectively. So there are 320 samples of actions in experiment. The model of camera is Panasonic AG-DVC180BMC DV (25 frames/second), and the resolution is 240 × 320.

B. Experimental results

Four methods were used to attain the recognition results and to compare the results with each other. They are MSF based SF-Tensor analysis(MSF SF-Tensor), SI based SF-Tensor analysis(SI SF-Tensor), MSF based K-means clustering(MSF Clustering) and SI based K-means clustering (SI Clustering) respectively.

There are 8 folds in variable views subspace, each of which has 4 × 10 images (SFI). Firstly 40 images in one fold

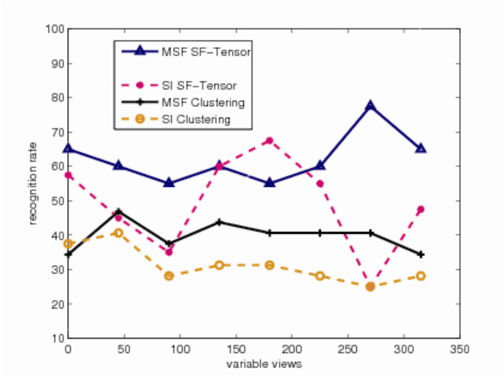


Figure 6. Multi-View Recognition Rate of Each Method

TABLE II. DIFFERENT PEOPLE RECOGNITION RATE OF EACH METHOD

	MSF SF-Tensor	*SI SF-Tensor*	*MSF Clustering*	*SI Clustering*
people1	81.25%	65.63%	37.50%	31.25%
people2	71.88%	78.13%	37.50%	31.25%
people3	62.50%	84.38%	43.75%	37.50%
people4	68.75%	62.50%	50.00%	31.25%
people5	68.75%	68.75%	40.63%	31.25%
people6	68.75%	65.63%	40.63%	34.38%
people7	68.75%	62.50%	40.63%	31.25%
people8	75.00%	68.75%	43.75%	37.50%
people9	81.25%	87.50%	53.13%	37.50%
people10	87.50%	90.63%	40.63%	40.63%

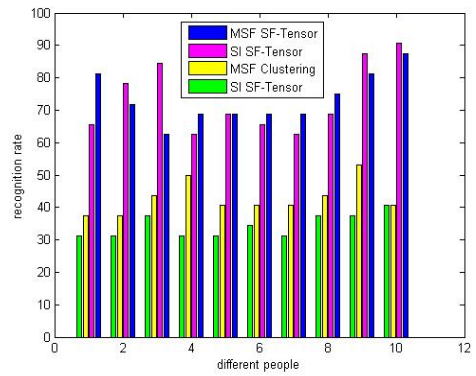


Figure 7. Different People Recognition Rate of Each Method

were selected randomly as the test set form D, while remaining 40 × 7 images of other folds were used as the training set, then the training progress was repeated 8 times From Table I and Fig. 6 it is easily observed that the MSF SF-Tensor has higher recognition rate than both the two K-means clustering methods. Meanwhile, considering the rate line of SI SF-Tensor, the highest point is above that of MSF SF-Tensor, while the two local lowest points are quite beneath, this is because the silhouettes are difficult to recognize for the similarity under the views of toward to and back away from the camera. The rate line of SI SF-Tensor.demonstrates that the recognition. rates change remarkably under multiple views.

TABLE III. MEAN RECOGNITION RATE BASED ON VIEW VS PEOPLE OF EACH METHOD

	MSF SF-Tensor	*SI SF-Tensor*	*MSF Clustering*	*SI Clustering*
people	73.44%	73.44%	42.81%	34.38%
view	63.13%	49.06%	39.84%	31.25%

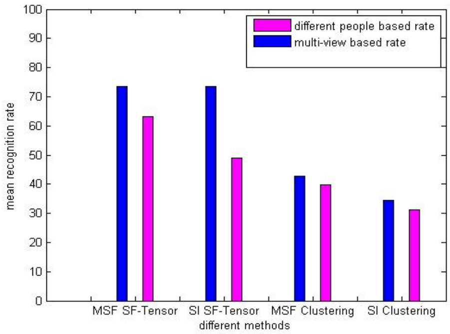


Figure 8. Mean Recognition Rate Based on View vs People of Each Method

There are 10 folds in variable people subspace, each of which has 4 × 8 images (SFI). Firstly 32 images in one fold were selected randomly as the test set form D, while remaining 32 × 9 images of other folds were used as the training set, then the training progress was repeated 10 times. From Table II and Fig. 7 it is obviously observed that both the SF-Tensor methods have higher recognition rate than the two K-means clustering methods. Meanwhile, from Fig.7 we can see that the distribution of recognition rates of each method is quite stable, which demonstrates that action recognition has not much relevance to different people.

Table III and Fig. 8 show the mean recognition rate of the four methods on condition of multi-view and different people separately. From Fig.8 it can be observed that both the SF-Tensor methods have higher recognition rate than the two K-means clustering methods on each condition. It is also obviously to see that the recognition rate of MSF SF-Tensor method is similar to SI SF-Tensor method considering different people, while the former is better than the latter under multi- view condition. Meanwhile the time-consuming of MSF SF- Tensor method is much less than SI SF-Tensor method because of the far less dimensions in the eigenSpace of the former.

V. CONCLUSION

A method of multi-view action recognition using tensor analysis based on multi-scale features(MSF) is proposed. The Serials-Frame based tensor (SF-Tensor) is decomposed to action subspace, view subspace, people subspace and Serials-Frame eigenSpace to recognize different actions constituted by a series of poses. The experiment results prove that the new method attain a good recognition rate in case of variable views and different people respectively. On the other hand the efficiency is highly improved by using multi-scale features.

ACKNOWLEDGMENT

This paper is supported by (1) the National Natural Science Foundation of China under Grant No. 60873146, 60973092, 60903097, (2) National High China under Grant No. 2007AAO4Z114, (3) Project of Science and Science (985 Engineering), (4) the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China, (5) National High Technology Research and Development Program of China under Grant No. 2009AA02Z307 (863), (6) the Key Laboratory for New Technology of Biology Recognition of Jilin Province No. 20082209, (7) the Third Phase Construction Project of ”211 Engineering” of Jilin University, (8) the Science and Technology Development Planning Project of Jilin Province (No.20080168).

REFERENCES

[1] B. Peng, G. Qian, and Y. Q. Ma, “Recognizing body poses using multilinear analysis and semi-supervised learning,” Pattern Recognition Letters, vol. 30, no. 14, pp. 1289–1294, 2009

[2] C. W. Ng and S. Ranganath, “Gesture recognition via pose classifica- tion,” in Proc. 15th International Conference on Pattern Recognition, vol. 3, Sep. 3–7, 2000, pp. 699–704

[3] Y.-T. Du, “Approach to human activity multi-scale analysis and recog- nition based on multi-layer dynamic bayesian network,” ACTA AUTO- MATICA SINICA, vol. Vol.35,No.3, pp. 225–232,2009.

[4] H.-T. C. Hsuan-Sheng Chen, “Human action recognition using star skeleton,” Santa Barbara, California, USA., vol. VSSN’06, pp. 171–178, October 27, 2006.

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, “Ghost: a human body part labeling system using silhouettes,” in Proc. Fourteenth International Conference on Pattern Recognition, vol. 1, Aug. 16–20, 1998, pp. 77–82.

[6] I. Cohen and H. Li, “Inference of human postures by classification of 3d human body shape,” in Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures AMFG, Oct. 17, 2003, pp. 74–81.

[7] S. Chinpanchana, “Semantic human action classification based on energy-action model,” IEEE, vol. 1/06, pp. 4244–0549, 2006.

[8] X. G. C. Tian, “Multi-view face recognition based on tensor subspace analysis and view manifold modeling,” in ElsevierB.V.Allrightsreserved. (72)3742-3750, 2009:

[9] C. Y. Chen, J. P. Zhang, and R. Fleischer, “Multilinear tensor-based non-parametric dimension reduction for gait recognition,” Advances In Biometrics, vol. 5558, pp. 1030–1039, 2009.

[10] L. Elden, “Matrix methods in data mining and pattern recognition.” SIAM, no. 4, 2007

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, “On the best rank-1 and rank-(r1,r2,...,r-n) approximation of higher-order tensors,” Siam Journal On Matrix Analysis and Applications, vol. 21, no. 4, pp. 1324–1342, 2000

[12] I. Stamos and L. Y. e. a. Liu, “Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes,” International Journal of Computer Vision, vol. 78, no. 2-3, pp.237–260, 2008

[13] X. G. Di. F.H.H, “Viewpoint insensitive pose representation for action recognition,” in Proc.AMDD, pp.143-152, 2006, pp. 143–152

[14] K. K. Sergios Theodoridis, Pattern Recognition, 2, Ed. China Machine Press, 2003

[15] M.-y. L. Zhen-hong Shang, “Real-time corner detection in binary image,” Journal of Image and Graphic,, vol. 10(3), pp. 295–230,2005