Human Action Silhouette Recognition Based on Tensor Analysis Using Synthetic Silhouette Data

Chan-Su Lee
Yeungnam University
South Korea
chansu@ynu.ac.kr

Yui Man Lui
Colorado State University
Fort Collins, CO USA
lui@cs.colostate.edu

Sung Yong Chun
Yeungnam University
South Korea
whiteyongi@ynu.ac.kr

Abstract

This paper presents performance evaluation of silhouette-based action recognition from virtual human silhouette(VIHASI), which is publicly available. We especially evaluate action recognition performance based on tensor analysis. We applied tensor analysis for the silhouette data itself after sampling the same number of silhouette frames from given sample frames. Geometric manifold embedding based on tangent bundle is applied and shows better performance compared with baseline performance.

1. Introduction

Understanding human activity is very useful in many applications; human activity recognition system can be used for visual surveillance systems [9], human-computer interactions based activity recognition [2], activity analysis from video analysis especially for sports videos, and so on. There has been a lot of research works related to activity recognition including gait analysis.

Recently, there are many research works related to human activity recognition [1], especially from video sequences. For better understanding of human activity, new features for activity recognition are proposed; modeling of the events in higher level are tried; even social relations are also considered [16] for better tracking and understanding of human activity.

For robust understanding of human activity in variant illumination, color, and background, silhouette data has frequently been used for activity recognition systems [4] and some silhouette databases are available for activity recognition especially gait recognition from CMU [3], Southampton [5], and USF [10]. However, there are few database which provides very accurate silhouette database with various activities, view variations, subjects. The virtual human action silhouette(ViHASi) data, which are publicly available, provides silhouette activity sequences generated using MotionBuilder graphics tools. The data base provides various view and different characters with 20 activities. This database can be used as a reference performance evaluation of activity recognition system based on silhouette sequences.

Recently there has been several approaches to recognize activities using tensor analysis. TensorGait is a model to apply tensor analysis for gait sequence from silhouette data [8]. Resampling based on circular manifold embedding is proposed to estimate the same frame number from gait sequence in different walking speed [7]. Tensor analysis based on within-class and between-class similarity is used for better recognition of individuals from gait sequences [15]. Based on tensor decomposition of activity sequence, tangent bundle on a Grassmann manifold is used for measurement of similarity from activity videos [12]. This is a kind of matrix manifold. More information about matrix manifold and its application in computer vision can be found from the review paper by Lui [11].

The paper applied for appearance video sequence activity recognition. In this paper, we applied the tangent bundle on a Grassmann manifold for activity recognition from silhouette data. We present a new and convenient silhouette normalization method based on given sequence without resampling. The proposed method shows better recognition performance compared to the baseline performance [13]. This is the first step to evaluate different tensor based approach for activity recognition from silhouette data.

2. Silhouette sequence processing

2.1. VIHASI database

The virtual human action silhouette(ViHASi) was generated for the purpose of evaluating various action recognition methods based on accurate silhouette data. The database provides 20 action sequences for 9 actors with up to 12 different viewpoints. 40 equally spaced different camera views with two sets of camera views are provided for evaluation of novel test cameras. We processed silhouette data based on the folder structure of the database. The order of the action index and the character index follows alphabetic orders as following tables shows.

Table 1. Action class names and its indexes

index	action name	index	action name
1	Collapse	11	KnockOutSpin
2	Granade	12	Punch
3	HangOnBar	13	Run
4	HeroDoorSlam	14	RunPullObject
5	HeroSmash	15	RunPushObject
6	JumpFromObject	16	RunTurn90Left
7	JumpGetOnBar	17	RunTurn90Right)
8	JumpOverObject	18	StandLookAround
9	Kicks	19	Walk
10	Knockout	20	WalkTurn180

In the case of character, we basically manage each folder as different subjects even though some of the sequences are from the same object in different views. Table 2 shows used character names and their index. When we test novel camera, however, we manage to combine the same object in different view as a single character data.

Table 2. Character names and its indexes

index	action name	index	action name
1	Gaulix1	7	Man4
2	Gaulix2*	8	MiaWorm
3	Humanoid1	9	Player
4	Humanoid2*	10	Wom1
5	Kids	11	Wom2
6	Man3		

*: same character with different viewing sets

2.2. Silhouette representation

We represent each shape instance as an implicit function $y(x)$ at each pixel x such that $y(x) = 0$ on the contour, $y(x) > 0$ inside the contour, and $y(x) < 0$ outside the contour. We use a signed-distance function such that

$$y(x) = \begin{cases} d_c(x) & x \text{ inside } c \\ 0 & x \text{ on } c \\ -d_c(x) & x \text{ outside } c \end{cases}$$

where the $d_c(x)$ is the distance to the closest point on the contour c with a positive sign inside the contour and a negative sign outside the contour. Such representation impose smoothness on the distance between shapes. Given such representation, the input shapes are points $\boldsymbol{y}_i \in \mathbb{R}^d, i = 1, \cdots, N$ where d is the same as the dimensionality of the input space and N is the number of points. Implicit function representation is typically used in level-set methods.

2.3. Spatial normalization:cropping and centering

Given silhouette data contains different number of frames in different actions and silhouettes are located in different places; the center of the silhouette moves due to actions and different view sets. In the action recognition based on silhouettes the normalization of silhouette location is very important; most action features based on silhouettes are dependent on the location of the silhouette in the image space. Therefore, we need to crop and normalize silhouette locations.

We first find the largest image blobs from a given silhouette data; in the given database, all the silhouette have well segmented and we can just use one image blobs as the largest image blobs. Then, we find the weight center based on foreground silhouette data. After finding the center of silhouette image, we need to crop the silhouette image around the center of the silhouette images. If we just use the center (x,y) points as the image center, the center will move around frequently during sequence of action; small movement of hand can cause change of a center location and moves the center of silhouette images. To overcome such a problem, we find a dominant axis of the silhouette data; in the most standing pose, the y axis (vertical axis) will be longer than x axis(horizontal axis), and the dominant axis is y axis in this case. After finding out the dominant axis, we normalize the center of the dominant axis moves to have equal distance into the blobs of extracted silhouette; we use center of the silhouette images for the other axis. Then the cropped image are normalized with equal sized images. In our experiments, the original images are normalized into 100 x 100 size images. Fig. 1 (b) shows cropped images after spatial normalization.

2.4. Temporal normalization: resampling to the same number of frames

The same number of image frames for each sequence is required to achieve tensor, or multilinear analysis [6, 14] of human action sequence from multi-dimensional array structure. Dynamic time warping can be used for temporal alignment of motion. Iterative temporal segmentation for best recognition tasks can be one of the ways to achieve good segmentation with good recognition performance [13]. Resampling based on manifold embedding can also be used [8]. Here we use relatively simple method based on known temporal segments.

For any given sequence with frame number N, we can find approximately equal interval M samples which is less than N by

$$\lceil \frac{N}{M}, \frac{2 \times N}{M}, \frac{3 \times N}{M}, \cdots, N \rceil \qquad (1)$$

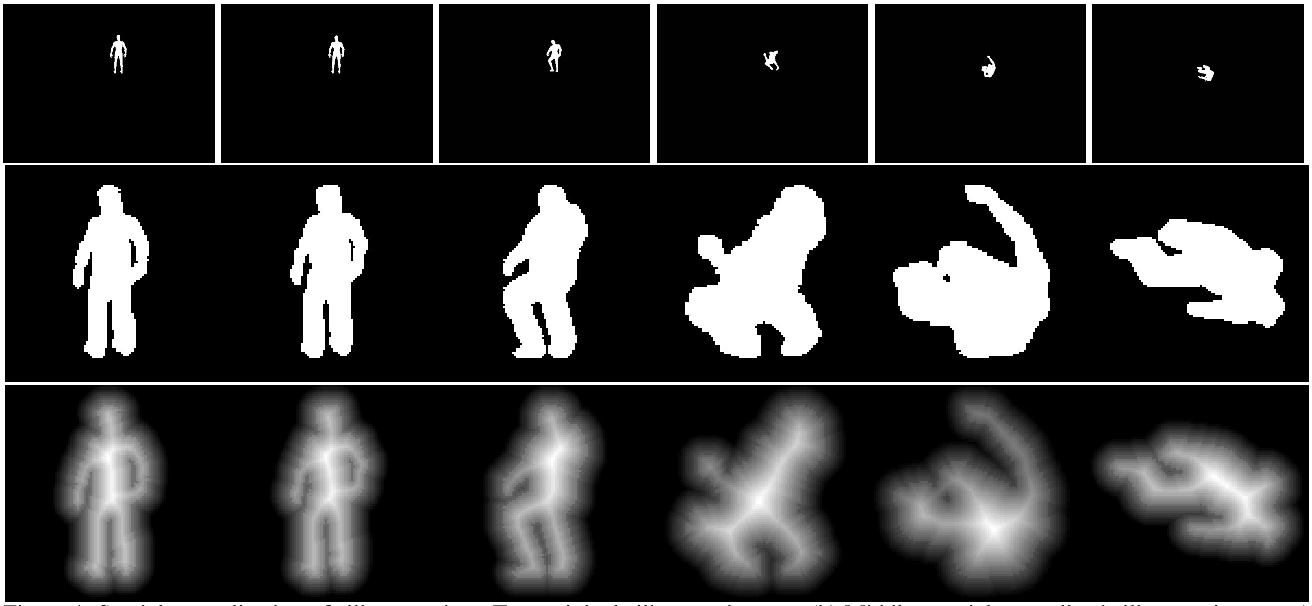

Figure 1. Spatial normalization of silhouette data. Top: original silhouette images. (b) Middle: spatial normalized silhouette images. (c) Bottom: signed distance representation of the cropped images.

, where $\lceil x \rceil$ means rounds the elements of x to the nearest integers towards infinity. This simple sampling method provides equal number of frame from a given sequence without resampling.

3. Tensor Based Human Activity Recognition

For the tensor-based human activity recognition from silhouette, we need to collect with multi-dimensional array. VIHASI database consists of 20 actions with at least 12 different views, 11 actors (here, we consider different view of the same actor as a different actor for convenience.), from 20 to 80 frames per sequences. We resampled the image size into 20×20 and each sequence with 20 samples by temporal normalization described in Sec. 2.4. This database can be collected into multi-dimensional array and can be factorized using higher-order singular value decomposition (HOSVD) [6] [14] as follows:

$$\boldsymbol{D} = \boldsymbol{Z} \times \boldsymbol{R} \times \boldsymbol{C} \times \boldsymbol{F} \times \boldsymbol{V} \times \boldsymbol{S} \times \boldsymbol{A} \qquad (2)$$

, where $\boldsymbol{D}$ is a data tensor with dimension $R_n \times C_n \times F_n \times V_n \times S_n \times A_n$, $\boldsymbol{Z}$ is a core tensor with the same dimension as data tensor, $\boldsymbol{R}$ is a row orthogonal basis with dimension $R_n \times R_n$, $\boldsymbol{C}$ is a column basis, $\boldsymbol{F}$ is a frame basis, $\boldsymbol{V}$ is a view basis, and $\boldsymbol{A}$ is a action basis. Fig. 2 shows examples of basis vector embedding in each subspace. We plotted in the 3 dimensional space using the first three elements for visualization purpose.

3.1. Tangent bundle for silhouette sequences

An action video can be represented by a third order tensor by $r \times c \times f$. Data tensor can be factorized to a set of

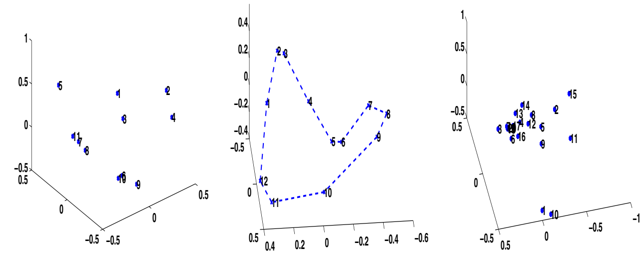

Figure 2. Examples of tensor subspace embedding. Left: style subspace. (b) Middle: view subspace. (c) Bottom: action subspace.

tangent spaces on a Grassmann manifold. In order to reside on the same Grassmann manifold, all factorized components are rescaled to have the same dimension. Distance on a Grassmann manifold can be estimated by means of tangent bundles.

The tangent bundle on a manifold is a set of tangent spaces to the manifold $\boldsymbol{M}$ defined as [12]:

$$\boldsymbol{T}(\boldsymbol{\mathcal{M}}) = \bigcup_{x \in \boldsymbol{\mathcal{M}}} \boldsymbol{T_x}\boldsymbol{\mathcal{M}}, \qquad (3)$$

where $\boldsymbol{T_x}\boldsymbol{\mathcal{M}}$ is the tangent space at the point x.

The logarithmic mapping, which is inverse of the exponential mapping, can be used to project a tangent space for associated points x on the manifold. Since the canonical metric on Grassmann manifold provides a means to measure the length of tangent vectors, the intrinsic distance for

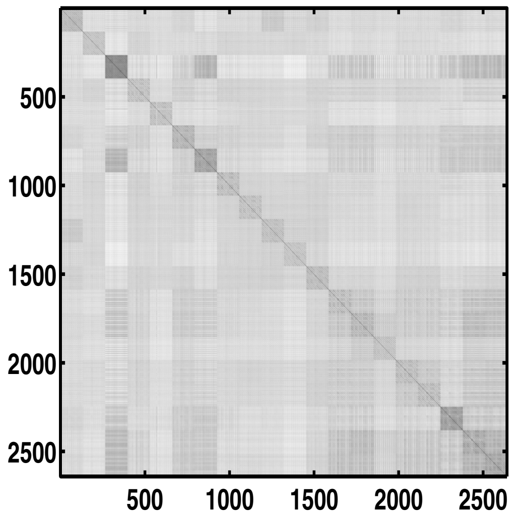

Figure 3. Pairwise distance matrix on a tangent space

the tangent bundles can be defined as follows:

$$\sum_{k=1}^{N} tr\{\delta_k^T \delta_k\}, \quad (4)$$

where the δ_k is the tangent vectors computed from the k^{th} tangent plane, and N is the order of the data tensor. The estimated intrinsic distance can be used for action classification.

4. Experiments

We followed the same first four protocols proposed in [13] for VIHASI to evaluate performance of action recognition from silhouette using tangent bundle. We represent each silhouette action sequence by $20 \times 20 \times 20$ a third order tensor; images are resampled to have size 20×20, and the temporally normalized to have 20 frames in each sequence. For all four cases, the proposed method shows better action recognition performance compared to the baseline results presented in [13].

4.1. Leave-One-Out Cross Validation

In this experiments, we estimate performance of action recognition in different sets. In the original experiment protocol, for each test action sequence, feature vectors in different camera view was removed. In our case, we represent feature vectors invariant to different camera views. The accuracy of the recognition is 99.47% with 2626 true positive(TP) and 14 false positive(FP). Table 3 shows confusion matrix for overall dataset. Fig. 3 shows pairwise distance based on our tangent bundle approaches.

4.2. Identical Train and Test Actors: Novel Test Cameras

For the novel test camera, we collected new data for the 2 actors (A1, A2) since previous data set contains common viewing angle. Here, 40 viewing angle for each subject is counted and trained for common view (V1, V3, V5, V7, V9, V11, V132, V15, V17, V19) and tested for the rest of camera views. Overall, the accuracy is 1196TP/(1196TP+4FP) = 99.67%, which is higher than the baseline accuracy 96.42%. Table 4 shows action recognition confusion matrix of this system.

4.3. Identical Train and Test Cameras: Novel Test Actors

In this novel actor test case, we trained action data for subject (A1 and A2) and tested with the rest of the actors (A3, A4, A5, A6, A7, A8, A9). Our approach achieved 98.98% accuracy which is higher than the baseline accuracy 98.04%. Detailed confusion matrix was given in Table 5.

4.4. Novel Test Actors: Novel Test Cameras

This novel test actors and novel test cameras is the most challenging case among the test four cases. We collected new data samples for this experiments as we need to dataset with (V2, V5, V7, V10, V12, V15, V17, V20) from actor (A1, A2) for training and we need test data for the rest of the data set except actor (A1 and A2). The performance reduced a little compared to other case; it shows good performance with recognition rate 98.87% (1661 TP /(19FP + 1661 TP)), whereas the baseline performance reduced into 89.82% with large drop of its performance. Overall confusion matrix was presented in Table 6.

5. Conclusions and Future Works

In this paper, we presented a novel approach for silhouette based action recognition based on Gassmann manifold embedding and distance measurement in this embedding space based on tangent bundle. In addition, we presented practical preprocessing and normalization method for extracted silhouette normalization, which will be very important for real silhouette dataset.

As the real video based data is available, we would like to apply proposed method to real dataset. In addition, we showed partial data of our experiment for VIHASI. We plan to complete our experiments with similar condition as the baseline case. Furthermore, we briefly explained other tensor based approach and did not have enough time to test and compare performance in different approach. We would like to further experiment for this VIHASI dataset using iterative tensor analysis and manifold embedding.

Acknowledgements: This research was financially supported by the Ministry of Education, Science Technol-

Table 3. Confusion matrix: leave-one-out evaluation result

Action	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
A1	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	1	0	130	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
A6	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0
A9	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0	0
A17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0	0
A18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	132	0	0
A19	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	0	120	6
A20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	132

Table 4. Confusion matrix: novel camera evaluation result

Action	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
A1	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	2	0	57	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
A6	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0
A9	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59	0	0	0	1	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0
A17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0
A18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0
A19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0
A20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60

ogy (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation and Basic Science Research Program through the National Research Fundataion of Korea(NRF) funded by the Ministry of Education, Science and Technology(2011-0005645)

References

[1] J. Aggarwal and M. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43:16:1–16:43, April 2011.

[2] J. Choi, Y.-I. Cho, T. Han, and H. S. Yang. A view-based real-time human action recognition system as an interface for human computer interaction. In *Proceedings of the 13th international conference on Virtual systems and multimedia*, VSMM'07, pages 112–120, Berlin, Heidelberg, 2008. Springer-Verlag.

[3] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report TR-01-18, Carnegie Mellon University, 2001.

[4] I. Haritaoglu, D. Harwood, and L. Davis. W4: Real-time surveillance of people and their activities. 22(8):809–830, August 2000.

[5] P. Huang, C. Haris, and M. Nixon. Recogising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[6] L. D. Lathauwer, B. de Moor, and J. Vandewalle. A multilinear singular value decomposiiton. *SIAM Journal On Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[7] C.-S. Lee and A. Elgammal. Homeomorphic manifold analysis: Learning decomposable generative models for human motion analysis. In *Workshop on Dynamical Vision*, 2005.

[8] C.-S. Lee and A. M. Elgammal. Dynamic shape style analysis: Bilinear and multilinear human identification with temporal normalization. *IJPRAI*, 24(7):1133–1157, 2010.

Table 5. Confusion matrix: leave-one-out evaluation result

Action	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
A1	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A6	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	82	0	0	0	0	0	0	0	0	0	0	1	1
A9	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0	0	0	0	1	1
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	0	0	0	5	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0
A17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0
A18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0
A19	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	76	7
A20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84

Table 6. Confusion matrix: novel view and actors

Action	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
A1	79	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	1
A2	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A6	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0
A9	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0	0	0	0	2	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0	0	2	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0
A17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0
A18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0
A19	0	0	0	0	0	1	0	0	0	0	0	0	5	1	0	0	0	0	74	3
A20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84

[9] M. Leo, T. D'Orazio, and P. Spagnolo. Human activity recognition for automatic visual surveillance of wide areas. In *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, VSSN '04, pages 124–130, New York, NY, USA, 2004. ACM.

[10] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. ICPR*, pages 211–214, 2004.

[11] Y. M. Lui. Advances in matrix manifolds for computer vision. *Image and Vision Computing*, 2011.

[12] Y. M. Lui and J. R. Beveridge. Tangent bundle for human action recognition. In *FG*, pages 97–102, 2011.

[13] H. Ragheb, S. Velastin, P. Remagnino, and T. Ellis. Vihasi: virtual human action silhouette data for the performance evaluation of silhouette-based action recognition methods. In *Proceeding of the 1st ACM workshop on Vision networks for behavior analysis*, VNBA '08, pages 77–84, New York, NY, USA, 2008. ACM.

[14] M. A. O. Vasilescu. Human motion signatures: Analysis, synthesis, recogntion. volume 3, pages 456–460, 2002.

[15] S. Wei, Y. Tian, and Y. Gao. Gait expression and recognition in the tensor space. In *Proceedings of International Conference Engineering and Computer Science(ICIECS)*, pages 1–4, 2009.

[16] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *Proceedings of IEEE Computer Vision and Pattern Recognition(CVPR)*, pages 1345–1352, 2011.