



HAL
open science

Trajectories Comparing Based on Matching and Distance Evaluation Within Stiefel and Grassmann Manifolds

Amani Elaoud, Walid Barhoumi, Hassen Drira, Ezzeddine Zagrouba

► **To cite this version:**

Amani Elaoud, Walid Barhoumi, Hassen Drira, Ezzeddine Zagrouba. Trajectories Comparing Based on Matching and Distance Evaluation Within Stiefel and Grassmann Manifolds. International Conference INFORMATION SYSTEMS IADIS, 2017, Lisbonne, Portugal. hal-01703958

HAL Id: hal-01703958

<https://hal.science/hal-01703958>

Submitted on 16 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRAJECTORIES COMPARING BASED ON MATCHING AND DISTANCE EVALUATION WITHIN STIEFEL AND GRASSMANN MANIFOLDS

Amani Elaoud¹, Walid Barhoumi^{1,2}, Hassen Drira³ and Ezzeddine Zagrouba¹

¹*Université de Tunis El Manar, Institut Supérieur d'Informatique, Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA), LR16ES06, Laboratoire de recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2 Rue Bayrouni, 2080 Ariana, Tunisia*

²*Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs, 2035 Tunis-Carthage, Tunisia*

³*IMT Lille Douai, CRISTAL (UMR CNRS 9189), Villeneuve-d'Ascq, France*

ABSTRACT

In this paper, we are interested in comparing human trajectories using skeleton information provided by a consumer RGB-D sensor. In fact, 3D human joints given by skeletons offer an important information for human motion analysis. In this context, the use of manifolds has grown considerably in the computer vision community in recent years. The main contribution of this study resides in working jointly with two manifolds. The matching of the trajectories is performed in Stiefel manifold and dissimilarity measure is carried out in Grassmann manifold. Indeed, trajectories of motions are provided by the projection in the Stiefel manifold. Then, the Stiefel distance is used within the dynamic time warping in order to define the appropriate matching between a reference trajectory and a test one. This allows avoiding that the rotation within the motion will be ignored, as it is the case with the Grassmann manifold. Then, the dissimilarity is evaluated using the Grassmann distance to compare motions while being invariant against rotation. Realized experiments on standard challenging datasets prove that the proposed method, for the comparison of human trajectories with different sizes, performs accurately compared to existing manifold-based methods.

KEYWORDS

Trajectories, human motion, manifolds, Grassmann, Stiefel, matching, Kinect, skeleton.

1. INTRODUCTION

Recognizing and analyzing human actions in videos is a challenging task with very complex actions in the real world (Barhoumi, 2014). Recently, skeleton-based human representations have been intensively studied and kept attracting an increasing attention. Some recent surveys, such as (Presti et Cascia, 2016), (Alashkar et al., 2016) and (Han et al., 2017), together overview this research area. The recent focus on skeleton representation has resulted in a variety of approaches, to conduct research on action recognition (Song et al., 2017) and modeling (Ding et Liu, 2017), tracking (Anjum et al., 2017) and re-identification (Wu et al., 2017). Indeed, the importance of the interpretation of human behavior from skeleton information was firstly reclaimed by (Shotton et al., 2011). We find success of skeleton information in action recognition such as (Vemulapalli et al., 2014) that use skeletal representation to prove the 3D geometric relationships of body parts in Lie group using rotations and translations in 3D space. Besides, (Du et al., 2015) propose a hierarchical recurrent neural network based on skeleton. The input skeleton is divided into five parts according to human physical structure, and a recurrent neural network is applied to classify the actions. Furthermore, a 3D action recognition method (Ke et al., 2017) with skeleton sequences is performed to jointly represent the sequence while incorporating spatial structural information. On the other hand, to deal with tracking issues, (Schubert et al., 2016) present automatic bone parameter estimation for skeleton tracking. The objective is to construct a rough hierarchical skeleton structure to optimize probabilistic skeleton tracking performance. Then, a method based on tracking the position of selected joints in human skeleton with depth videos recorded using Kinect camera is proposed in (Anjum et al., 2017). More recently,

(Wang et al., 2014) extract the skeleton of the human body using the 3D joint positions that are generated via skeleton tracking from the depth map sequence. (Meng et al., 2016) study the human gait from skeleton data using inter-joints distances as spatio-temporal features. Various skeleton-based representations are also used for human re-identification. For instance, (Munaro et al., 2014a) use skeleton information to build a descriptor which can then be classified with an SVM (Support Vector Machine) classifier for one-shot person re-identification through a consumer depth sensor. Recently, deep learning approaches have been customized for re-identification. For instance, (Barbosa et al., 2017) propose a framework through a classification based on CNN architecture.

In recent years, there is a growing attention for the techniques reformulating computer vision problems over non-Euclidean spaces, such as Riemannian manifolds. The study of these manifolds has important consequences for applications such as dynamic textures (Miao et al., 2017), human activity modeling and recognition (Feng et al., 2017), face recognition (Cai et al., 2016) and shape analysis (Amor et al., 2016). For instance, (Turaga et al., 2008) use Stiefel and Grassmann manifolds for analysis applications in computer vision, such as activity recognition, face recognition and shape classification. (Lui et Beveridge, 2011) represent videos as a tangent bundle on a Grassmann manifold for action classification. Videos are expressed as third order tensors and factorized to a set of tangent spaces. Then, tangent vectors are computed between elements on a Grassmann manifold. More recently, (Slama et al., 2014) use Grassmann manifold and model motion for activity recognition from depth sequences. In another work, (Michalczuk et al., 2017) present an application of Grassmann manifold to the evaluation of the human stability based on movements in hip, knee and ankle joints. In general, the use of manifold transforms informations on trajectories, that is why temporal modeling and time-warping solutions have attracted an important interest to compare sequences, trajectories or motions. An emerging solution is to compare two sequences with different sizes. The same person can do the same action with different rates. Indeed, the same activity can greatly decrease recognition performance, if some frames are ignored. The invariance to the temporal rate of execution of action helps to provide an accurate recognition. (Turaga et Chellappa, 2009) propose as a model a Time-Varying Linear Dynamic System (TV-LDS) and Dynamic Time Warping (DTW) for the problem of modeling and recognizing complex activities in time-varying dynamic. (Abdelkader et al., 2011) use the DTW algorithm in order to solve the temporal alignment and to match a given time series with the optimal non-linear warping function. (Amor et al., 2016) propose to use human shape evolution of skeletons as trajectories on Kendall shape manifold via smoothed transported square-root vector fields (TSRVFs). More recently, (Devanne et al., 2016) analyze the trajectory of motion and consider shape variations within a Riemannian manifold to learn step models. They propose to analyze their shape by using the SRVF, where each shape trajectory is viewed as an element of the shape space.

As best we know, all the existing manifold-based methods chose one manifold that is then used for matching as well as for distance evaluation. In this paper, we are interested in using the DTW algorithm, the popular sequence alignment and matching method, especially for two unregulated sequences. Indeed, it can resolve the problem of temporal alignment and it measures the similarity between sequences varied in time in order to be capable of providing superior performance with size variance. The main idea of this work is the projection of a set of videos. Each frame will be a point on a manifold, and construct a trajectory of motion to lay these points. Then, Stiefel distance d_S is applied for the matching in the DTW and then the Grassmann distance d_G is applied for the couples resulting from the matching. We chose to apply in this paper this idea in the framework of human motion analysis, and more precisely for people re-identification which is an important and challenging task. It is helpful for many applications in various domains, like entertainment, medicine and surveillance. As many-to-one matching, between a reference trajectory and a test one, is often used, we opted for the Stiefel manifold in order to avoid invariance against rotation. Indeed, if a rotation-invariant manifold is used during the matching, which is totally spatial, one frame in the reference video can be wrongly matched with many frames in the test video, notably when the filmed persons rotate on themselves. However, as the temporal information is also considered during the assessment of the similarity between two motions, we opted for the Grassmann manifold which is invariant to rotation.

The remainder of this paper is structured as follows. Proposed method is discussed in Section 2. Section 3 presents experimental results and numerical evaluations. Concluding remarks and ideas for future works are given in Section 4.

2. PROPOSED METHOD

We study the problem of evaluating similarity of human motions using 3D videos generated by an RGBD sensor, especially the Kinect. By representing human body as dynamical trajectories, we study the evolution of their sequences of skeletons as trajectories on the non-linear manifolds. A manifold is a topological space locally similar to the Euclidean space. In this work, we are interested in two relevant manifolds: Stiefel and Grassmann. On the one hand, Stiefel manifold is the set of k -dimensional orthonormal bases in R^n , where $k < n$, represented by $V(R^n)$. We obtain an $n \times k$ matrix Y , such that $Y^t Y = I_K$. To define a Stiefel metric between two elements of this manifold, the Frobenius norm is widely used. In fact, given two elements P and P' of Stiefel manifold, the Stiefel distance is defined by:

$$d_S(P, P') = \|P - P'\|_F \quad (1)$$

where $\|\cdot\|_F$ is the standard Frobenius norm. It is equal to the square root of the matrix trace of $M \cdot M^t$, defined by:

$$\|M\|_F = \sqrt{\text{tr}(MM^t)} \quad (2)$$

On the other hand, Grassmann manifold $G(R^n)$ is a quotient space of $V(R^n)$, represented by $V(R^n)/SO(K)$, where $SO(K)$ is the orthogonal group of dimension k . Two points on a Grassmann manifold are equivalent if one can be mapped into the other one by a rotation matrix: a $k \times k$ orthogonal matrix (Edelman et al., 1998). Then, to define a Grassmann distance, the length of the shortest curve θ connecting the subspaces S and S' generated by point P and the best matched point P^* and it is computed according to:

$$d_G(S, S') = \sum_{i=1}^n \theta_i^2 \quad (3)$$

The proposed method ignores the RGB channel and uses only the skeleton data. In fact, given a skeleton sequence, composed of a set of ordered frames that shows a person in motion, we try to measure similarity between trajectories. We start by analyzing shapes of human skeletons, and their temporal evolutions. Since human skeletons are characterized by sets of registered points (or landmarks), each skeleton is characterized by N joint points obtained by the Kinect sensor. Note that N is equal to 20 (*resp.* 25) in the case of Kinect-Version1 (*resp.* Kinect-Version2) as was described by (Elaoud et al., 2017). Then, each skeleton is projected in the Stiefel manifold. The idea behind using the Stiefel manifold representation is that each skeleton of 3D information can be as a point representation which results the appearance of the trajectories containing these points. Thus, for two videos V and V' with different sizes, we obtain two trajectories T and T' . Then, we try to find the most similar point in T with each point in T' while using the distance d_S , what results in defining the best matched couples. For example, as we can see in Figure 1, P_1 in T is matched with P'_2 in T' , P_2 in T is matched with P'_4 in T' ... and P_n in T is matched P'_m with in T' . The matching is continued until the end of the trajectory and we keep the best matched point P_j^* defined by:

$$P_j^* = \text{Minarg}_{1 \leq j \leq n} d_S(P_i, P'_j) \quad (4)$$

The next step consists to calculate the distance d_G between the matched couples. To do this, the distance d_G is computed between $(P_1, P_2^*), (P_2, P_4^*), \dots$ and (P_n, P_m^*) . Thus, the dissimilarity D is the sum of Grassmann distances $d_G(P_1, P), d_G(P_2, P_4^*), \dots$ and $d_G(P_n, P_m^*)$.

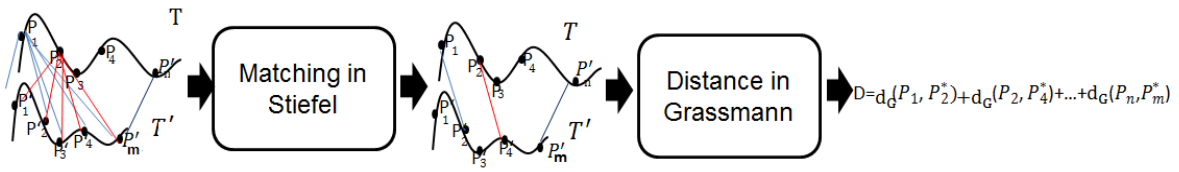


Figure 1. The outline of the proposed method

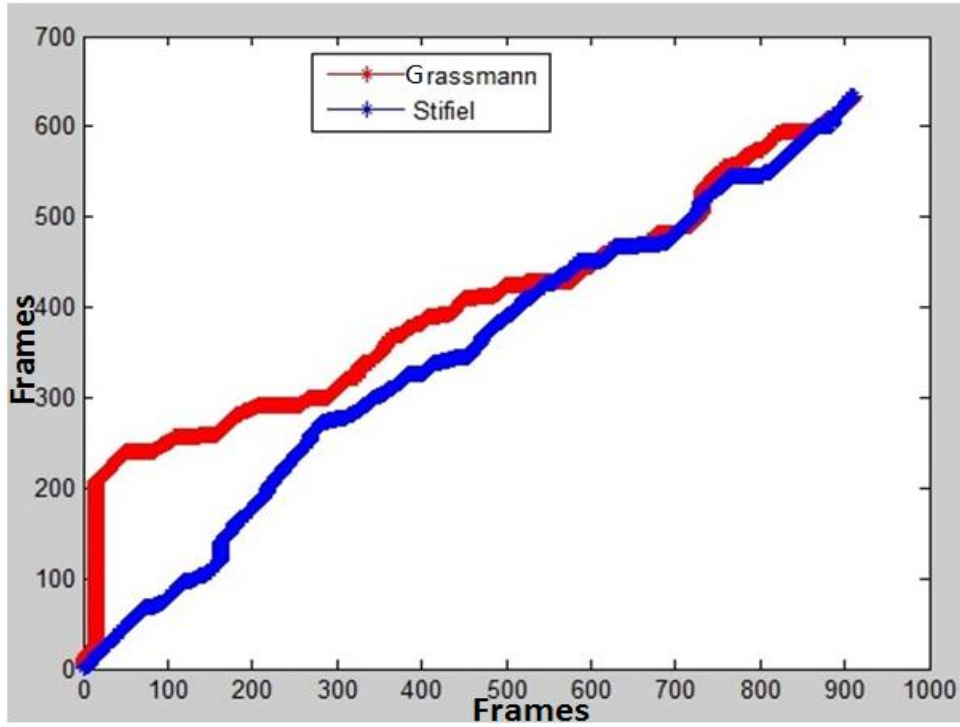


Figure 2. The matching in Stiefel and Grassmann manifolds

As described previously, we explore two different manifolds (Stiefel and Grassmann) using two different metrics in order to calculate dissimilarity between motions. This can be helpful to analyze actions, classifications and re-identification issues. Using the mathematical framework described in the previous sections and the nearest neighbor as a classifier, we can evaluate the dynamic of skeletal data. Thus, in our case, a point on a Stiefel manifold is an orthonormal matrix of size 20×3 that can be viewed as 3-dimensional subspace of R^{20} . By applying the DTW with the Stiefel distance, we obtain the nearest point. The matching in Stiefel is used to guarantee the best matching since Grassmann is known by the ignorance of rotation. As we can see in Figure 2, by ignoring the rotation with Grassmann, we can match many frames (200 frames) with the same frame, if the difference in motion is just a rotation. Then, Grassmann distance $d_G(2)$ is applied for all couples of landmarks already matched (Figure 2). Finally, the dissimilarity D is the sum of matched couples all along trajectories as described in what follows:

```

Input: Two trajectories  $T$  and  $T'$  using skeleton information

Output: Dissimilarity  $D$  between  $T$  and  $T'$ 

Method: Matching of the trajectories in Stiefel manifold and
dissimilarity measurement in Grassmann manifold

BEGIN
D ← 0
/* Browse the trajectory T of size n */
for i=1 to n
/* Identify the point  $P_j^*$  in  $T'$  that matches best the point  $P_i$  in  $T$  */
Tab ← Null /* Fill array with matched values with  $P_i$  */
min ← 0
index ← 0 /* The index of the min value */
/* Browse the trajectory  $T'$  of size m */
    for j=1 to m
        Tab[j] ←  $d_S(P_i, P_j')$ 
        if Tab[j] < min
            min ← Tab[j]
            index ← j
        end
    end
end
D ← D +  $d_G(P_i, P_{index}^*)$ 
End
END

```

3. EXPERIMENTS

This work addresses the problem of identifying a person (e.g. in a company, in an airport...) and can reduce the problem of scanning device, waiting in line or remembering user passwords. This method can be used to identify a person by the way of walking (comparison trajectories of motion (Elaoud et al., 2017)). In this section, we present some experimental illustrations, in order to evaluate the proposed method on challenging datasets, such as the BIWI-Lab RGBD-ID dataset (Munaro et al., 2014b). It is a specified RGB-D dataset for re-identification tasks that aims to re-identify different persons in different locations with different appearances. It is divided on 50 training sequences and 56 testing sequences of 50 different people. In the training sets, people repeat the same action in front of a Kinect sensor, like a rotation around the vertical axis, several head movements and two walks towards the camera. 28 people, different of the 50 ones that are present in the training video, have been recorded also in two testing sequences each. These testing sets have been recorded in a different day and in a different location with respect to the training dataset, therefore most subjects have different clothes. Every person has a Still sequence and a Walking sequence in the testing set. For the first set, the Still videos, every person is slightly performing motion in place and for the second set, Walking sequences, every person moves with two walks frontally and two other walks diagonally with

respect to the Kinect.

For evaluation purposes, we compute Cumulative Matching Curves (CMC) (Gray et Tao, 2008), which are commonly used for analyzing the re-identification performances. In Figure 3.a and Figure 3.b, we report the CMC curves on Still BIWI-Lab RGBD-ID and Walk BIWI-Lab RGBD-ID datasets, respectively. The reported curves are based on Stiefel distance d_S , Grassmann distance d_G and matching Stiefel with Grassmann (noted d_M)-based classifications. As shown in Figure 3.a (for the set Still BIWI-Lab RGBD-ID), the distance d_G and the distance d_M -based classifications performs better results than the one based on d_S in most of the cases. The results reported using the distance d_G and the distance d_M are similar in general.

In fact, the recognition rate starts in rank 2 for both distances with a value of 3.57%. Then, for rank 4, the distance d_M -based classification outperforms the distance d_G -based classification with a value of 14.28% against 10.71%. For the others ranks, the curves are alternated, such as in rank 20 to 33 the distance d_G -based classification outperforms distance d_M -based classifications but in rank 35 to 50 the distance d_M -based classification outperforms the one based on the distance d_G . Especially, the first distance that reaches 100% for the recognition rate is the d_M distance in rank 48 against 92.85% for d_G . As shown in Figure 3.b (for the set Walk BIWI-Lab RGBD-ID datasets), the distance d_G -based and the distance d_M -based classifications perform better results than the one based on d_S . The recognition rate at rank 3 is equal to 7.14% for both d_G and d_M -based classifications. Then, one can see along the CMC curves that the d_M -based classification is better than the ones based on the distance d_G with a value of 10.71% at rank 4 against 7.14% for the classification based on d_G . The results for both distances are close in general, such as the d_M -based classification outperforms the d_G -based classification in rank 12 to rank 23 but the d_G -based classification is the best in ranks 23 to 30. A finer comparison detailed these results and the maximum values are presented in Table 1 and Table 2. This leads us to conclude that it cannot be the best measure in all cases but it has performance for more than half of results. Then, the matching can exist as a measure and can give good results. Thus, as best as we know, it is the first time that this unsupervised method is proposed and can have an important role. As re-identification with Kinect is still recent, there are not many previous approaches in this area to make a complete comparison. To the best of our knowledge, limited efforts have been spent on Kinect-based person re-identification (Munaro et al., 2014a, 2014b). The proposed method in (Munaro et al., 2014a) and (Munaro et al., 2014b) is based on anthropometric measurements that are calculated from the skeleton data and not on the gait. The gait extracted from skeleton for re-identification is more challenging task and has not been sufficiently investigated. We argue that the proposed method is among first efforts to re-identify person based on their gaits using skeleton data.

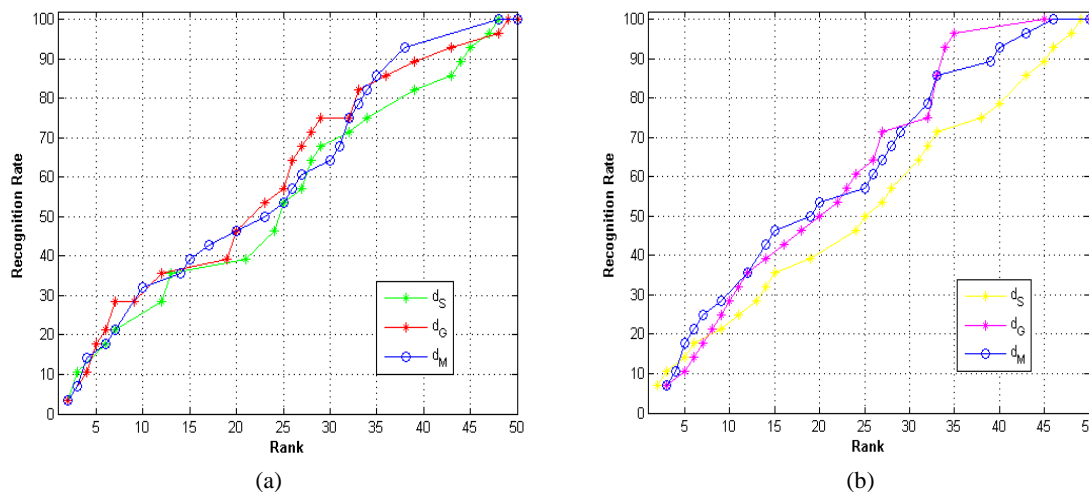


Figure 3. Re-identification accuracy using Still BIWI-Lab RGBD-ID dataset and Walk BIWI-Lab RGBD-ID dataset

Table 1. Comparison of re-identification results using Still BIWI-Lab RGBD-ID dataset

Rank	D_G	D_M
2-4	10.71%	14.27%
4-9	28.57%	27%
9-11	31%	32.14%
11-14	35.14%	34%
14-20	39.28%	42.85%
20-32	75%	68%
32-34	82%	82.14%
34-48	96.42%	100%
50	100%	100%

Table 2. Comparison of re-identification results using Walk BIWI-Lab RGBD-ID dataset

Rank	D_G	D_M
1-12	32.14%	35.71%
12-23	64.28%	67.85%
24-30	75%	71.42%
30-33	77%	78.57%
33-44	98%	97%
46	100%	100%

4. CONCLUSION

We have presented in this paper an accurate method for comparing human motions using evolutions of skeletons. The proposed method is based on metrics that are independent of the color information and change appearance. Indeed, we used the Riemannian geometry to process skeletal data and their temporal evolutions with a classification performed by closest neighbor. The main contribution of the proposed method resides in working conjointly with two manifolds: Stiefel and Grassmann. The matching of the trajectories is realized in Stiefel manifold and the dissimilarity assessment is performed in Grassmann manifold. This permits to benefit of to two manifolds. This work is tested in the field of re-identification, but it can be applied for others applications such as human action recognition or tracking. Although that the produced matching by our work is not the best in all cases, notably when compared with supervised methods, but this work introduces for the first time the possibility of using two different manifolds for the challenging stages of matching and distance computing. Future work will be focused on more manifolds and different types of distances to be able to provide a good measure of dissimilarity.

REFERENCES

- Abdelkader, M.F. et al., 2011. Silhouette-based gesture and action recognition via modeling trajectories on Riemannian shape manifolds. *Computer Vision and Image Understanding*, 115(3), pp. 439–455.
- Alashkar, T. et al., 2016. A Grassmann framework for 4D facial shape analysis. *Pattern Recognition*, 57, pp. 21–30.
- Amor, B.B., Su, J. & Srivastava, A., 2016. Action Recognition Using Rate-Invariant Analysis of Skeletal Shape Trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1), pp. 1–13.
- Anjum, M.L., Rosa, S. & Bona, B., 2017. Tracking a Subset of Skeleton Joints: An Effective Approach towards Complex Human Activity Recognition. *Journal of Robotics*, 2017, pp. 1–8.
- Barbosa, I. B. et al., 2017. Looking beyond appearances: Synthetic training data for deep CNNs in re-identification. *arXiv preprint arXiv:1701.03153*.
- Barhouni, W., 2014. Detection of highly articulated moving objects by using co-segmentation with application to athletic video sequences. *Signal, Image and Video Processing*, 9(7), pp. 1705–1715.
- Cai, Z. et al., 2016. RGB-D datasets using microsoft kinect or similar sensors: a survey. *Multimedia Tools and Applications*, 76(3), pp. 4313–4355.

- Devanne, M. et al., 2016. Learning shape variations of motion trajectories for gait analysis. 2016 23rd International Conference on Pattern Recognition (ICPR).
- Ding, W. and Liu, K., 2017. Learning linear dynamical systems with high-order Tensor data for skeleton based action recognition. arXiv preprint arXiv:1701.03869.
- Du, Y., Wang, W. & Wang, L., 2015. Hierarchical recurrent neural network for skeleton based action recognition. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Edelman, A., Arias, T.A. & Smith, S.T., 1998. The Geometry of Algorithms with Orthogonality Constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2), pp.303–353.
- Elaoud, A. et al., 2017. Analysis of Skeletal Shape Trajectories for Person Re-Identification. *International Conference Advanced Concepts for Intelligent Vision Systems*.
- Feng, Y., Lu, B. & Zhang, D., 2017. Multiscale singular value manifold for rotating machinery fault diagnosis. *Journal of Mechanical Science and Technology*, 31(1), pp. 99–109.
- Gray, D. & Tao, H., Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. *Lecture Notes in Computer Science Computer Vision – ECCV 2008*, pp. 262–275.
- Han, F. et al., 2017. Space-time representation of people based on 3D skeletal data: A review. *Computer Vision and Image Understanding*, 158, pp. 85–105.
- Ke, Q. et al. 2017. A New Representation of Skeleton Sequences for 3D Action Recognition. arXiv preprint arXiv:1703.03492.
- Lui, Y.M. & Beveridge, J.R., 2011. Tangent bundle for human action recognition. *Face and Gesture 2011*.
- Meng, M. et al., 2016. Detection of Abnormal Gait from Skeleton Data. *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*.
- Michalczuk, A. et al., 2017. Evaluation of human dynamic balance in Grassmann manifold.
- Miao, J. et al., 2017. Manifold Regularized Slow Feature Analysis for Dynamic Texture Recognition. arXiv preprint arXiv:1706.03015.
- Munaro, M. et al., 2014. One-Shot Person Re-identification with a Consumer Depth Camera. *Person Re-Identification*, pp. 161–181.
1. Munaro, M. et al., 2014. 3D reconstruction of freely moving persons for re-identification with a depth sensor. 2014 IEEE International Conference on Robotics and Automation (ICRA).
- Presti, L.L. & Cascia, M.L., 2016. 3D skeleton-based human action classification: A survey. *Pattern Recognition*, 53, pp. 130–147.
- Schubert, T. et al., 2016. Automatic bone parameter estimation for skeleton tracking in optical motion capture. 2016 IEEE International Conference on Robotics and Automation (ICRA).
- Shotton, J. et al., 2011. Real-time human pose recognition in parts from single depth images. *Cvpr 2011*.
- Slama, R., Wannous, H. & Daoudi, M., 2014. Grassmannian Representation of Motion Depth for 3D Human Gesture and Action Recognition. 2014 22nd International Conference on Pattern Recognition.
- Song, S. et al., 2017. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. In. *AAAI*, pp. 4263–4270.
- Turaga, P. & Chellappa, R., 2009. Locally time-invariant models of human activities using trajectories on the grassmannian. 2009 IEEE Conference on Computer Vision and Pattern Recognition.
- Turaga, P., Veeraraghavan, A. & Chellappa, R., 2008. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. 2008 IEEE Conference on Computer Vision and Pattern Recognition.
- Vemulapalli, R., Arrate, F. & Chellappa, R., 2014. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. 2014 IEEE Conference on Computer Vision and Pattern Recognition.
- Wang, J., Liu, Z. & Wu, Y., 2014. Learning Actionlet Ensemble for 3D Human Action Recognition. *SpringerBriefs in Computer Science Human Action Recognition with Depth Cameras*, pp. 11–40.
- Wu, A., Zheng, W.-S. & Lai, J.-H., 2017. Robust Depth-based Person Re-identification. *IEEE Transactions on Image Processing*, pp. 2588–2603.