Computer Simulation Modeling for Human Motion

Ali Yu

College of Physical Education, Jiangxi Science and Technology Normal University, Nanchang, 330013, China

Abstract
In the background of progressive completed sparse representation theory, sparse representation of signal has gradually aroused increasing attentions from scholars and has been widely used in various fields. At the same time, the digital media industry, represented by three-dimensional films and games, rises gradually and the computer animation technology develops greatly, which has become a hotspot between scholars. Due to the unique space structure and time structure in human motion capture data, to simply use existing sparse representation model theory is difficult for the motion capture data analysis and processing. Therefore, how to effectively use the sparse representation theory to make a better motion capture data modeling, analyze and process the human motion capture data so as to achieve better results has become a significant issue. A semi supervised learning algorithm based on sparse representation is proposed in this paper to fully excavate the potential rules in the motion, obtain the Mahalanobis distance measurement. On these grounds, the logical similarity between two motions is judged to conduct the motion modeling and simulation. Experimental results have given prominence to the algorithm advantages.

Key words: SPARSE REPRESENTATION, HUMAN MOTION, CAPTURE DATA

Introduction

Sparse representation is the key problem to delve in signal processing, data analysis and other fields. Using relevant algorithms can acquire signals or data in a redundant dictionary with a concise representation, that is, most of the coefficient in the vector is zero and there are only a small amount of non-zero coefficients. Those non-zero coefficients and the corresponding atoms typically reveal the essential characteristics of signals or data as well as their internal structures, which can simplify the subsequent processing tasks (1). With the rapid development of information technology, data acquisition method has become more and more abundant. The growing expansion of the large amounts of data have appeared in many fields, such as video and audio data, human motion capture data, life information data, web data, and so on. There is an urgent need for a simple, flexible and adaptive expression. On the other hand, with the rapid development of motion capture technology, the human motion capture data has become a new type of media data with rapid growth. The technology of motion capture has also been widely used in 3D movies, video games, human-computer interaction and sports training, etc. [2]. However, the cost of data acquisition by motion capture equipment is very high, and we need to hire a professional actor to do the long-playing data acquisition. And after the data acquisition, we need professionals to make the data collection of clips and after-treatment. Therefore, the research on the reuse of motion capture data is always a hot research in graphics animation field. Because sports data has a strong constitutive property with high data dimension and is a kind of time series data with high pause rate, it is difficult for ordinary tools to conduct the corresponding analysis and processing [3]. Sparse representation modeling is the exact answer to dealing with these problems. How to reasonably use the sparse representation model for motion capture data modeling and how to generate a 3D human animation with richer types and more natural representation by data processing method from the existing motion capture data is an important problem to be solved urgently.

In recent years, relevant researchers of human motion have also introduced the sparse representation model to the analysis of the mining motion data. The high dimension of motion capture data causes inconvenience to the motion capture data analysis and processing. Since human motion capture data has a strong constitutive property, and the motion data with similar type often have similar internal structure, so we can often describe them with a small amount of feature parameters. Taking advantage of the sparse representation to make human motion capture data modeling is a good idea. It can both simplify the capture data, and well reveal the internal motion characteristics and laws of the motion data, which can provide a better condition for the motion capture data analysis and processing.

At present, there are also some relevant scholars who have applied the sparse representation into the human motion capture data processing. For example, in the bibliography [8], it regards each human motion sequence as a set of the joint time-domain signal. Due to the inherent sparseness of each signal, we can obtain the common base of each motion sequence and the sparse representation of each joint timing signal, and then the sparse coefficients can be used for the classification of motions, motion retrieval, and motion data compression, which have already gained a very good effect. In the bibliography, we propose a learning algorithm based on the sparse representation-- Regularized Distance Metric Learning with Sparse Representation (RDSR). Based on the geometrical feature descriptor, this algorithm finally obtains the distance metric matrix through semi supervised learning. This distance measurement matrix can integrate the relationship between tags and unsupervised data perfectly, which can be used as the code of transition fragments of the epidemic and applied to the motion retrieval. However, most of the existing researches have directly applied the sparse representation model into the motion capture data in accordance with the usage modes of images or signals. Although it can achieve some certain effects, if we only use the unique motion characteristics to model the sparse representation, the adaptability of the algorithm will be greatly reduced, and it will be more difficult for the practical application. Therefore, how to reasonably use the sparse representation model to make the motion capture data modeling and how to generate a 3D human animation with richer types and more natural representation by data processing method from the existing motion capture data is an important research objective in this paper.

The semi supervised learning algorithm based on sparse representation in this paper shall be used for similarity measurement of human motion capture data. The retrieval technique of human motion is a necessary link in motion data management and reuse process. It is difficult for Euclidean distance to measure the logical similarity between two motions. Consequently, the semi supervised learning algorithm based on sparse representation is proposed in this paper to excavate logical similarity between motions by training marked motions, fully excavate the potential rules between motions by training unmarked motions and obtain the Mahalanobis distance measurement. On these grounds, the logical similarity between two motions is judged to conduct motion retrieval. The method can obtain higher inquiry accuracy and be applied in automatic retrieval without any manual intervention.

Feature Extraction

Human motions captured by different motion capture devices have different human skeleton information. In order to adapt to the different data structure, this section only uses 15 joints to extract the feature without damaging key information. These 15 joints are: head, neck, root, Right/Left shoulder, Right/Left elbow, hand of Right/Left, Right/Left hip, Right/Left knee, Right/Left foot. Usually, two motion data in time is + K, which needs to use the dynamic time warping (DTW) to align the two data in the dimension of time.

However, this would cause a certain information loss and some unnecessary interference. A similarity motion retrieval method based on the human body posture coding is proposed in the bibliography. We can encode the motion data according to the feature representation, divide continuous pauses with same encoding into equivalent segments, establish inverted index with dimensional foundation of the equivalent segment to retrieve candidate motions and gain the final inspection results by coding matches. However, the final result adopt DTW algorithm to calculate the similarity of similar motions and inquire samples, which causes information losses and lead to unnecessary disturbances to the result. Based on the JRD, the bibliography proposes the relative distance variance of the joint (VJRD) as the feature, and obtains a better result. Variance represents the fluctuation range of each JRD in the average value. From the visual point of view, if they are the two similar logical motions, then the motions of these joints are the same, and these characteristics will reflect on the value of the VJRD. VJRD in the trade and transportation can avoid the alignment operation of the DTW as well. However, VJRD only includes the information of the joint distance. For the complex human motion data, the amount of information contained is not enough. So we use the GPD which is full of features to replace the JRD as the feature.

This section uses VGPD as a feature of a motion sequence. For the motion sequence M, it can be formally represented as a frame $M = \{F_1, F_2, \cdots, F_7\}$, which contains the motion data of T frame $F_t\left(1 \le t \le T\right)$, and represents each one of the frames. For each frame, we need to calculate 1683 features. The sequence of the tectonic sequences are GPD, and when the sequences finally gain the GPD, it can be calculated simply by the formula:

$$VGPD_M\left(p\right) = \frac{1}{T}\sum_{t=1}^{T}\left(nGPD_M\left(t,p\right) - \overline{nGPD_M\left(p\right)}\right)^2$$

VGPD can not only describe the logical relationship between motions, but also cannot be affected by the motion sequence length. It simplifies the intermediate process and improves the efficiency of retrieval, which is conducive to the automation of motion retrieval.

Semi Supervised Distance Learning Based on Sparse Representation

Human motion database can be expressed as $X = \{x_1, x_2 \cdots, x_t, \cdots, x_N\}$ formally, in which the number of motion sequences in the database is represented by N. As for the labeled data, we used $y_u \in \{0,1\}$ to indicate whether i and j motion is the same category of motion The Euclidean distance between any two motion sequences is expressed as: $d\left(x_i, x_j\right) = \sqrt{\left(x_i - x_j\right)^T\left(x_i - x_j\right)}$

The advantage of Euclidean distance similarity measurement is the higher computational efficiency, but the disadvantage is that the calculation process ignores the semantic interpretation of the motion data features. This measurement method cannot gain the same perception as the human beings do, which means the logically similar motions are not similar in the motion values. The logically similar motion distances obtained are often larger, while the non-similar motion distances are much smaller. Consequently, a similarity computation function identical to the feature description of motion data is obtained by learning the distance measurement, namely $d\left(x_i, x_j\right)_M$, where M is the inverse of the covariance matrix. In the new transformation space, the logically similar motion distance is small while the dissimilar motion distance is large. So the new distance measurement formula can be expressed as: $d\left(x_i, x_j\right)_M = \sqrt{\left(x_i - x_j\right)^T M\left(x_i - x_j\right)}$

We should make sure that M is a symmetric and semi positive matrix. So we can write M as. Then the distance measurement formula can be expressed as: $d\left(x_i, x_j\right)_M = \sqrt{\left(x_i - x_j\right)^T W^T W\left(x_i - x_j\right)}$

The purpose of this section is to get the best W through the optimization learning, so we need to define a series of loss functions to achieve the goal. By introducing a series of loss function, the similarity measure criterion of the frames is obtained by optimization in the bibliography. Similar to the literature, we need to make full use of the label information in the motion data firstly. The different part from the literature is that when the two motion data belong to the same type, we regard that they are similar, which means that they are semantically similar, namely, the value distance of logically similar motions should be as small as possible. So we need to guarantee that the distance between the same type of motion data should be small, and the distance between different types should be large. We use $E_{\sin ular}$ to represent the distance squared between the same types of motion data. $y_{ij} = 1$represents that I motion and j motion are of same type, namely, $S_{\sin ular} = \sum\left(x_i - x_j\right)\left(x_i - x_j\right)^T$. Here we need to ensure that the distance between the same types of motion data is as small as possible, which means that $E_{\sin ular}$ is as small as possible. We can also use $E_{\sin ular}$ to represent the distance squared between the same types of motion data:

$$E_{\sin ular}$$

$$= \sum_{y_u=1} \left(x_i - x_j\right)^T WW^T \left(x_i - x_j\right)$$

$$= Tr\left(\sum_{y_y=1}\left(W^T\left(x_i - x_j\right)\left(x_i - x_j\right)^T W\right)\right)$$

$$= Tr\left(W^T S_{stmilar} W\right)$$

On the other hand, in the practical application, the number of labeled training is often very small, and there would be a lot of unlabeled data. As a result, a semi supervised learning method is proposed in this section. We hold that any motion sequence can be performed by other motion sequences and that the motion data of the same type should be represented by a similar representation method. Suppose that $X = \left[x_1, x_2, \cdots, x_N\right]$ indicates a matrix composed of motion data of the same type. Then each column Xi represents a motion sequence, $\boxed{X}$ stands for a matrix composed of unlabeled motion data, and each column represents a movement sequence. Then we can get: $X \approx \boxed{X}A$

Each labeled motion data can be represented by an unlabeled motion data, a_1 represents the reconstructed coefficient of each motion sequence X. Due to massive unlabeled motion data, it is usually the case that only a part of the motion data can be reconstructed well. At the same time, if the motion data belong to the same type, the reconstruction of these motion data should be similar. That is to say, using the same unlabeled motion data can well reconstruct all the X motion data in X. The above two aspects of the characteristics are similar to the—Group-lasso. So in this section, we introduce it to solve the value of A. Group-lasso can render some of the rows in X as 0, which means that motion data of same type is reconstructed with the same batch of unlabeled data, which is in line with the human intuition. $A = \arg\min\left\|X - \boxed{X}A\right\|_F^2 + \lambda\left\|A\right\|_{2,1}$

The distance between the original motion data and the reconstructed motion data should also be small, and this information needs to be preserved in the study to denote the distance square between them as well:

$$E_{ai}$$

$$= \sum_{i=1}\left(x_i - \boxed{X}a_1\right)^T WW^T\left(x_1 - \boxed{X}a_1\right)$$

$$= Tr\left(W^T\left(\sum_{i=1}\left(x_1 - \boxed{X}a_1\right)\left(x_1 - \boxed{X}a_1\right)^T\right)W\right)$$

$$= Tr\left(W^T X\left(I_N - A\right)\left(I_N - A\right)^T X^T W\right)$$

$$= Tr\left(W^T XS_{oi} W\right)$$

But we should pay attention to that a large number of unknown types of data are contained in the unlabeled data. These motion data is very important in the same way. We can use them in a direct way, but the different part is that we should get rid of the data we used before. Similarly, in the unlabeled data matrix $\boxed{X}$, suppose that each motion sequence Xi can be linearly reconstructed by other motion sequences, so the coefficient of reconstruction in the motion sequences should be sparse. We can use the following formula to solve this problem.

$$a_i = \arg\min\left\|x_i - \boxed{X}a_i\right\|_2^2 + \lambda\left\|a_i\right\|_1$$

Similarly, we use E_{sr} to indicate the distance between them:

$$E_{sr}$$

$$= \sum_{i=1}\left(x_i - \boxed{X}a_i\right)^T WW^T\left(x_i - \boxed{X}a_i\right)$$

$$= Tr\left(W^T\left(\sum_{i=1}\left(x_i - \boxed{X}a_i\right)\left(x_i - \boxed{X}a_i\right)^T\right)W\right)$$

$$= Tr\left(W^T X\left(I_N - A\right)\left(I_N - A\right)^T X^T W\right)$$

$$= Tr\left(W^T XS_{sr} W\right)$$

It can be written as:

$$W^* = \arg\max\frac{Tr\left(W^T AW\right)}{Tr\left(W^T BW\right)}$$

When we get a new inquiry sample, we can use the Mahalanobis distance measurement to obtain the distance of other motion data in the database, and then put these distances in an ascending order and thus get the search result.

Experiments and Evaluation

(A) Human motion capture experiment

The labeled data in this experiment is obtained from the HDM05 motion capture database, containing 3634 different motion segments. All these motion segments are motions of a single type, containing 52 different kinds of motions. The total database size is 720MB. To verify the validity of the method in this section, we consider a part of the labeled data as the unlabeled data. First of all, we select 90% of the motions from the database, half of which is used as the training sample set and the other half as the searching database for the test. In the training sample set, we select 40 types of the motion data, of which 40% is used for supervised learning, the remaining 60% and other 12 types are used as the unlabeled motion data for training. The remaining 10% of the database is retrieved as the test sample set. The TopN strategy is used to measure the retrieval quality. When retrieving the accuracy of different motions, the total number of samples in the corresponding type of the database is N. Repeat the above procedure 10 times to carry out the cross validation.

At the same time, the curve of the retrieval accuracy and recall rate is given. In order to verify the validity of the methods in this section, we select some typical methods to make comparison. For walking motions, such as walking and lime walking, we can divide them into many types. When the motion of the training database does not contain the lime walking, it will be difficult for us to use the methods in the bibliography to retrieve the motion data. It is also very difficult to solve the problem by methods proposed in literatures. In this section, we obey the same internal rules and use the Group-lasso to explore the relationship between the labeled motion and the unlabeled motion for the similarity metric learning, which can be applied in various occasions. The curve in Fig. 1 also shows that the retrieval accuracy and recall rate of the proposed method in this section have both reached a relatively high level.

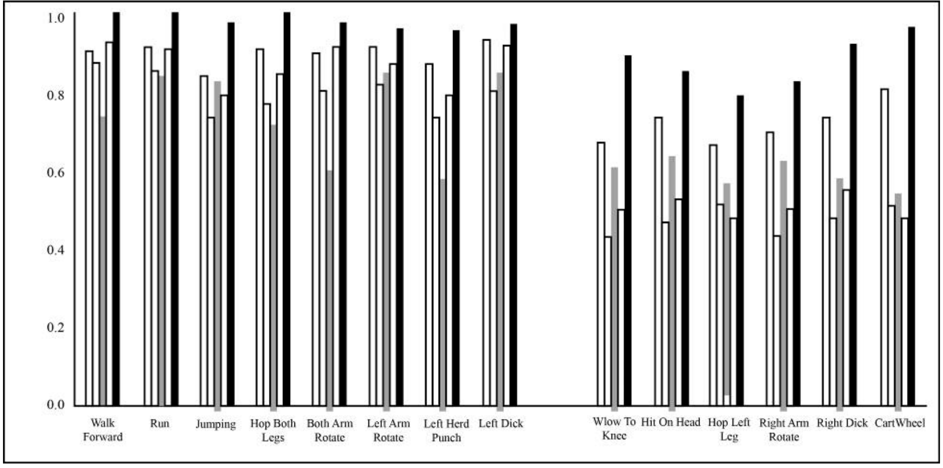

Figure 1. Retrieval accuracy of the motion data under TopN strategy

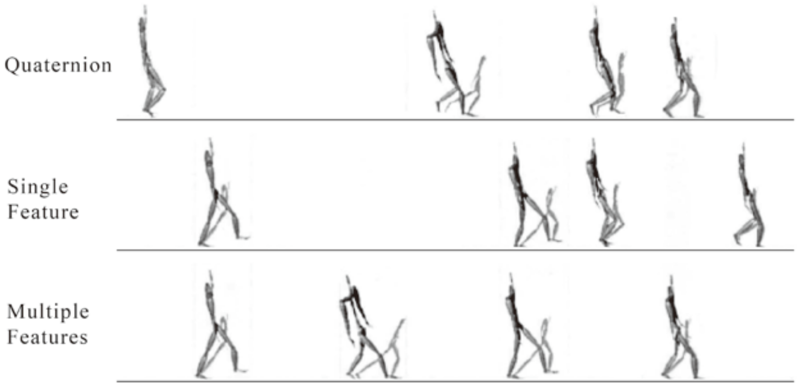

Figure 2. Key frame extraction of the different walking feature in 4 periods

Figure 2 shows the key frame sequences of different features for the walking motion in four periods. Different walking motions have different features. Each row of the key frames is sorted from left to right according to the time sequence. In this section, the effect of motion parameters adjustment is reflected by comparing the motion parameters of the synthetic motion before and after the adjustment. We select walking motion and running motion to demonstrate the composite results.

Figure 3 shows the average motion of 100 walking motions in the database. The figure 4 is the plan view of the average motion. By adjusting the low dimensional space parameters, users can control the motions on the basis of the average motion, and then a new walking motion sequence can be synthesized. The walking direction change is the most obvious, so the result of the experiment mainly shows the adjustment to the direction.

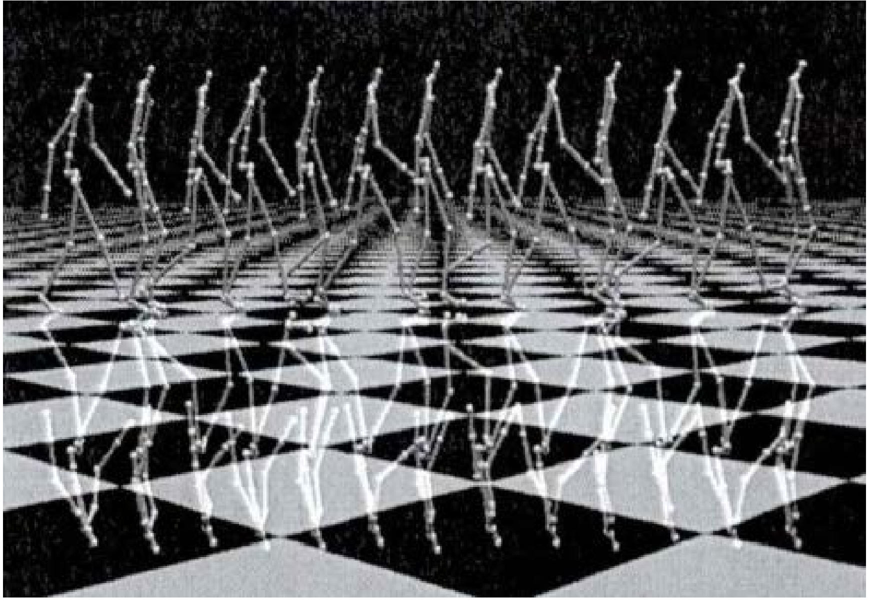

Figure 3. Average motion of the walking motion

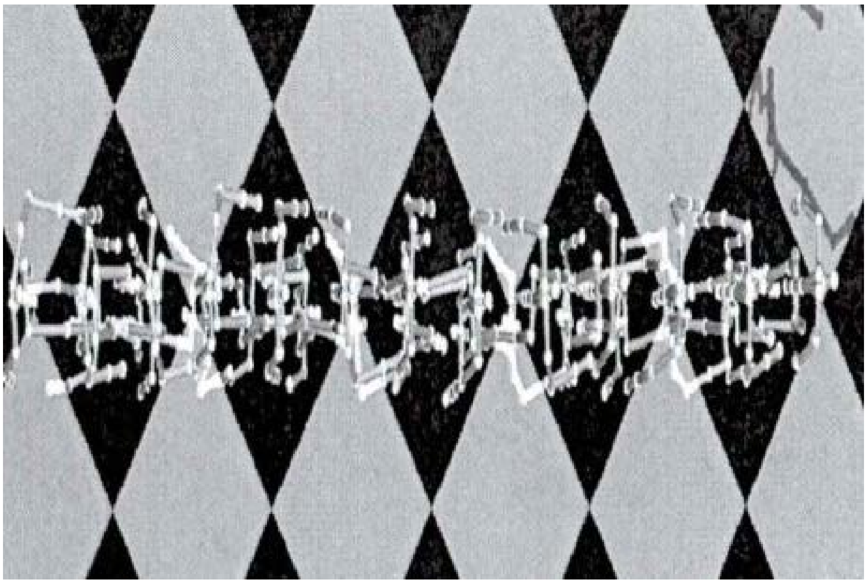

Figure 4. Plan view of the average walking motion

(B) Key frame extraction experiment of Human motion capture data

The experiment data in this paragraph are from CMU Graphics Lab Mo-cap Database and HDM05 Database. To verify the effectiveness of model in this paragraph, motion document with different types and properties are selected in the experiment, including single-period, multi-period circulatory, complex and multi-type motions. The motion types include shadowboxing, kicking, walking, running, jumping, cartwheel and shoot. To compare the quality of key frames extracted by different methods, 15 persons who have no animation experience are selected to judge the quality of key frames. The key frames vary from people, for the purpose of eliminating these differences, statistical analysis is conducted to the key frames picked by such 15 persons. Due to the higher blunt rate in motion capture data and similar with adjacent frames, key blunt obtained is a scope, as shown in 4.4, the correct key blunt shall be obtained when the extracted key frames fall in green area.

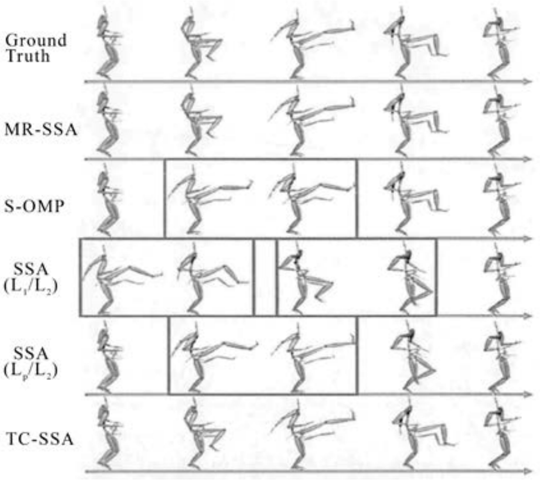

Figure 5. Extraction Results of Key Frames

Conclusions

This paper contains a lot of meaningful research questions. It has focused on the motion retrieval, the key frame extraction and the synthesis to make an in-depth study, reasonably established the sparse representation model, and put forward the concrete solutions, which have effectively solved the practical problems in this field. Quantifying internal loads resulting from muscles and external forces is necessary to treat and understand movement pathologies. However, current simulation methods fall short of accurately reproducing observed human performance when being compared to observations made in motion experiments. It is a minimal requirement that models closely reproduce external observations before considering them to answer questions about severity of possible treatments and a disorder. Given this underlying requirement, a dynamical motion-tracking method has been developed to leverage experimental observations directly to guide forward simulations and provide simulations with higher accuracy.

References

1. Sun Y. (2010) *Image sparse representation and its application in image processing inverse problems*. Nanjing University of Science and Technology.
2. Wang Z., Xia H. (2009), Synthesis of Virtual Human Progress. *China science*, 45(05), p.p.483-498.
3. Yang J, Ganesh A. (2009) Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), p.p.210-227.
4. Yang J, Huang J, Ma T.Y. (2010) Image Super-resolution via Sparse Representation. *IEEE Transactions on Image Processing*, 19(11), p.p.2861-2873.
5. Shen B, Hu W, Zhang Y. (2009). Image in Painting via Sparse Representation, *IEEE International Conference on Speech and Signal Processing*, Kobe, Japan, p.p.697-700.
6. Li H, Li C, Yi H. (2011) Adaptive Feature Extraction using Sparse Coding for Machinery Fault Diagnosis. *Mechanical Systems and Signal Processing*, 25(2), p.p.558-574.
7. Yi L, Fermuller C, Aloimonos Y, Hui J. (2010) Learning shift-invariant sparse representation of actions. *IEEE Conference on Computer Vision and Pattern Recognition*, p.p. 2630-2637.
8. Chen C, Zhuang Y, Nie F. (2011). Learning a 3D Human Pose Distance Metric from Geometric Pose Descriptor. *IEEE Transactions on Visualization and Computer Graphics*, 17(11), p.p.1676-1689.