An Improved DTW Method for Human Behavior Recognition

Shengnan Li[a], Yan Ma[b, *], Hui Huang[c], and Shunbao Li[d]

Collage of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China

[a]lsn_199310@163.com, [b]ma-yan@shnu.edu.cn, [c]huanghui@shnu.edu.cn, [d]lsb@shnu.edu.cn

*The corresponding author

Keywords: Human Behavior Recognition; Kinect; DTW; Curve Fitting; Behavior Similarity

Abstract: Human behavior recognition has been widely used in computer vision. However, the accuracy and stability of the recognition algorithm will be affected by the different data acquisition objects and object behavior. To solve this problem, an improved human behavior similarity measure method based on improved dynamic time warping (DTW) is proposed. First, the data of human joints is acquired by Kinect. Then, the path is corrected with DTW algorithm according to the motion distance between the motion to be tested and template. Next, the corrected path is further adjusted via curve fitting. Finally, the behavior similarity between them is obtained by the adjusted path. The experimental results show that the performance of the proposed method is more stable and accurate.

1. Foreword

Human behavior recognition [1] is one of the hot research directions of computer vision. This research field includes many disciplines: signal processing, pattern recognition, machine learning and so on. Its research results are widely used in human-computer interaction. Intelligent video surveillance and other fields.

The design of a good human behavior recognition algorithm should pay attention to three aspects: (1) the data is easy to collect and takes up less space; (2) the accuracy of the algorithm is high; and (3) the running time of the algorithm is short. In previous studies, researchers mainly use two-dimensional image [2] and video sequence [3 - 5]. Under the influence of light and shadow, two-dimensional image and video are easy to make target recognition unstable. With the advent of big data's era, the traditional methods have some problems, such as insufficient data storage space and weak expansibility of algorithms.

Microsoft launched Kinect somatosensory devices, can effectively avoid the impact of environmental factors such as lighting. The data obtained by contactless data has high accuracy, small memory occupation [6] and low equipment cost, which provides a better data source for human behavior recognition research. In recent years, Kinect has been widely used in the field of human behavior recognition [7-10]. Zhu GM [11] uses Kinect to study continuous human action recognition, and proposes an online classification method based on likelihood probability to improve the detection of the starting and ending points of behavior.

In view of these problems, this paper proposes to improve DTW to better combine Kinect action recognition. The main work of this paper is as follows: (1) to change the limit of DTW slope and (2) to modify DTW regular path by curve fitting.

2. Data Acquisition

Using Microsoft Kinect somatosensory device, using depth information, can accurately identify the human skeleton node. There are 25 knots, each of which is numbered and located as shown in figure 1.

  DOI: 10.25236/isrme.2019.033

$$J = (j_0, j_1, j_2 \ldots, j_{23}, j_{24})$$

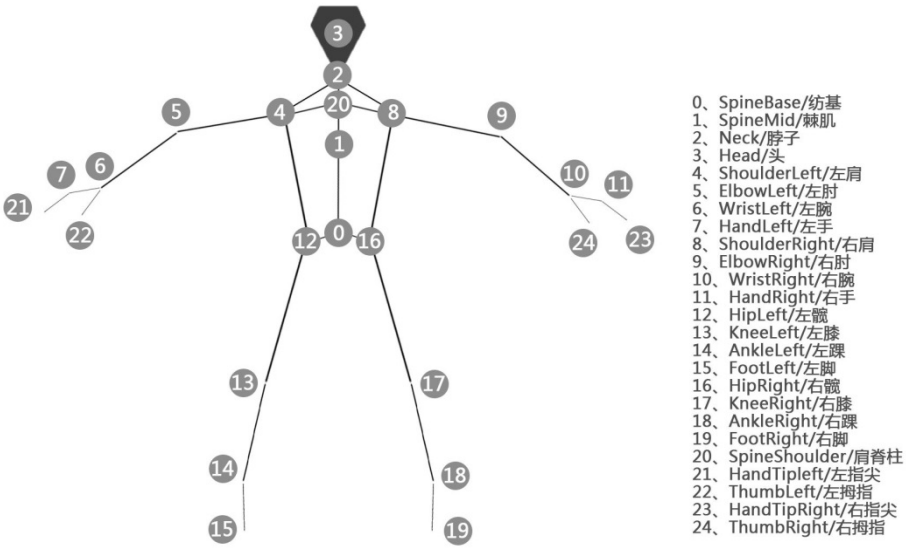


Figure. 1 Kinect Human body Node

The node information obtained by Kinect is depth image data, which is first converted to three-dimensional spatial coordinates. The transformation process can call the mapping function provided by Kinect for Windows SDK. The node information is recorded as J, It is expressed as

The location information for each node consists of three dimensions, recorded as j_i Expressed as $j_i = (x_i, y_i, z_i),\ i \in [0,24]$ The 25 nodes of the template action information are recorded as t, which is represented as: $t = (j_0^t, j_1^t, j_2^t, \ldots, j_{23}^t, j_{24}^t)$.The 25 nodes of the action information to be detected, recorded as a, are represented as: $a = (j_0^a, j_1^a, j_2^a, \ldots, j_{23}^a, j_{24}^a)$.

The template behavior information is recorded as T, which includes m action information. Where ti represents the 25 node information of the action information of item (i+1)th in the template information, T is represented as $T = (t_0, t_1, t_2, \ldots, t_{m-1})$.

The behavior information to be detected is recorded as A and includes a total of n action information. Among them, the information of 25 key nodes of the action information of Article (I 1) of the information to be detected is represented by the information of the action information of Article (I 1) of the information to be detected, Where ai represents 25 joint point information of the (i-1) action information in the information to be detected, A is represented as: $A = (a_0, a_1, a_2, \ldots, a_{n-1})$.

3. Behavior Similarity Algorithm

The behavior similarity algorithm calculates the similarity between the test behavior and the template behavior. Firstly, the distance between each action in the template behavior and each action in the behavior to be detected is calculated, then the behavior action to be detected is corresponding to the template behavior action by DTW algorithm, and then the path generated by DTW is modified, and finally the behavior similarity is calculated.

3.1. Distance between Actions.

Calculate the similarity of action. First calculate the distance between the corresponding nodes in t and a $d_q(j_q^t, j_q^a)$

$$d_q(j_q^t, j_q^a) = \sqrt[2]{(x_q^t - x_q^a)^2 + (y_q^t - y_q^a)^2 + (z_q^t - z_q^a)^2},\ q \in [0,24]$$

Then add up the distances to get the aggregate distanc $D(t,a)$:

$$D(t,a) = \sum_{q=0}^{24} (d_q(j_q^t, j_q^a)$$

3.2. DTW Algorithm.

State Time Warping DTW (Dynamic Time Warping) Is A Nonlinear Dynamic Warping Technique Which Combines distance measure calculation with time warping method. The template behavior information T contains m action information, and the behavior information A to be detected contains n action information. The distance information between the actions to be tested A and the template behavior T is mapped to the first quadrant of the two-dimensional coordinate system, the horizontal axis represents n actions of the behavior A to be tested, and the longitudinal axis represents m actions of the template behavior T. A total of m*n grid coordinate points are formed. Coordinate point(c , r) represents the distance difference between the C action information in A and the r action information in T $D(a_c,\ t_r)$, Of which $c \epsilon [0, n-1], r \epsilon [0, m-1]$.

In order for each action in A to correspond to a unique action information in T, you need to find it from the grid point optimal path from coordinate point (0,0) to point (n-1,m-1), Minimize the distance between T and A. Let the optimal path of the strip be composed of n coordinate points $P_i(i, r_i)$,which $i \in [0, n-1]$, $r_i \in [0, m-1]$, r_i indicates that the ith action in A corresponds to the r_i action in T, that is, $P_i(i r_i)$ represents the distance between the ith action in A and the r_i action in T $D(i,\ r_i)$.

The path adjusted by the above dynamic time should meet the following requirements:

1) Monotonicity: $r_i \geq r_{i-1}$, $0 \leq r_i \leq m-1$;

2) The starting point $P_0(0,\ r_0)$must be $(0,\ 0)$, the destination $P_{n-1}(n-1,\ r_{n-1})$ must be $(n-1, m-1)$

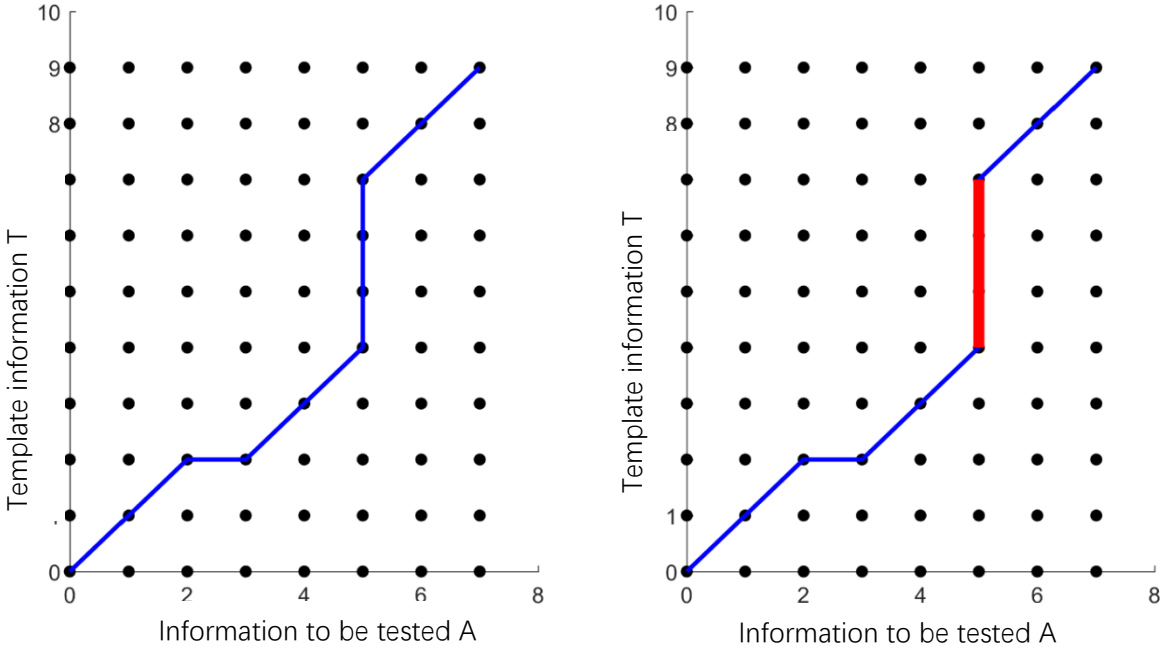
3) Step size constraints: $r_i - r_{i-1} \geq 0$.

There are two situations in the DTW algorithm: when the n=m , calculate the punctuation point *P*（*1,1*）,*P*（*2,2*）,…,*P*（*n,n*）in turn, corresponding distance values can be added to the sum. When n ≠ m, the distance values of the coordinate points corresponding to the coordinate points on the DTW planning path and the template information should be calculated, and the sum of the values should be calculated. According to the actual situation of Kinect, even n = m has the problem of fast and slow action in continuous behavior, so the corresponding distance sum can not be calculated directly in turn, and the route should be planned.

The matching calculation of all paths and all the nodes in these paths leads to a great deal of computation, and the amount of computation can be greatly reduced by dynamic programming.

3.3. DTW Path Correction Algorithm.

After the path matches, DTW plans an optimal path, as illustrated by the path (a). In figure 2 According to the DTW principle, on this path, a transverse coordinate corresponds to a unique vertical coordinate, but the longitudinal coordinate may correspond to multiple transverse coordinates, for example, the bold part of (b) in figure 2.



(a) DTW Common path Graph (b) DTW path Multi-valued to one-valued Annotation

Figure. 2 DTW Planning path

This will cause the problem of behavior similarity calculation, that is, which action to be tested should match the template. Therefore, the least square curve fitting path is proposed and the selected path of DTW is adaptively modified. The principle is that if one action to be tested corresponds to multiple template actions, the fitting curve passes through as the final result. There are two special cases: (1) if the fitting curve does not pass through in this case, the nearest point from the fitting curve is taken as the final result; (2) if the starting point of the behavior to be tested is not the same as that of the template behavior, in this case, there must be no fitting curve at the beginning or end point, and the distance between the fitting curve and the fitting curve is not the same. If the distance is too large, the starting or ending point of the line segment is redetermined to be the starting or ending point of curve fitting. In this way, the action to be tested can only correspond to a template action.

The path slope of the original DTW is controlled between 0.5 and 2, but it cannot be done in human behavior. Because the action group included in the behavior, according to the speed of the behavior of the parties and the frequency of equipment collection, it is very common to cause the gap in the number of actions between the action groups is too large.

3.4. Behavioral Similarity Calculation.

According to the modified DTW path, the distance value of each point on the path is calculated, and the distance between T and A is obtained by summing up and taking the mean value $D(T,A)$,:

$$D(T,A) = \frac{\sum_{u=0}^{n} D(a_{c_u},\ t_{r_u})}{n}, n \in [1, +\infty]$$

Calculation of behavior similarity the calculation of reference action similarity is introduced into the Sigmoid function, α is a preset error factor, Similarity calculation between template behavior T and behavior A to be detected:

$$S(T,A) = \left(\frac{1}{1+\ e^{D(T,A)/\alpha}}\right) * 2, D \in (0,+\infty), S \in (0,1)$$

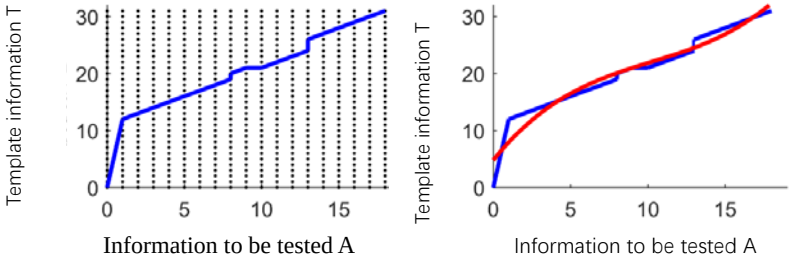
For detecting that the behavior to be tested and the template behavior start or end point is different, still need to deal with one step. Assuming that the starting point of the behavior to be tested corresponds to the first action of the template in the modified route after curve fitting, and the end point of the behavior to be tested corresponds to the b action of the template, then the similarity between the template behavior T and the behavior A to be detected is as follows:

$$S(T,A) = S(T,A) * \frac{b-a+1}{m}, a \in [0,m-1], b \in [0,m-1]$$

At this point, the similarity between the test behavior and the template behavior is obtained.

3.5. Behavior Similarity Result Analysis.

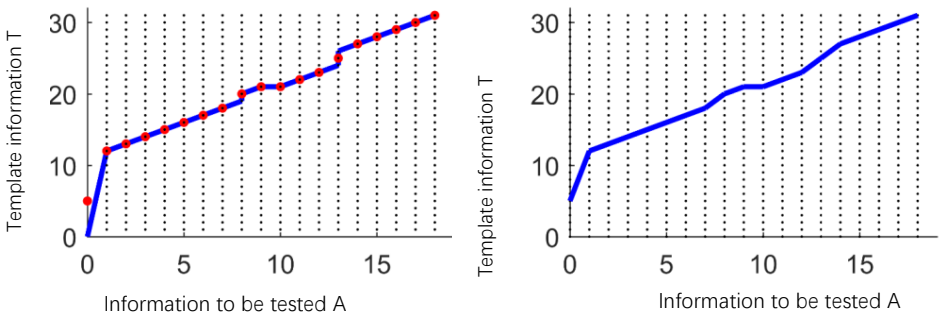
The salute action is selected as an example, in which the template behavior information has 32 action information, and the behavior information to be tested has 19 action information. In figure 3, (a) is the path of DTW matching between the action to be detected and the template action, and you can see that there is an obvious vertical, which is that multiple r values correspond to a c value. In figure 3, (b) is the path after fitting the DTW curve, where red is the fitting path and blue is the DTW path.



(a) DTW Planning path (b) DTW path Curve fitting

Figure. 3 DTW path fitting process

In fig. 4, (a) selects the correction point of DTW path according to curve fitting. It can be seen that the action matching information starts and ends neatly, and the middle multi-value corresponds well to a single value. In figure 4, (b) is the final path after modifying the original optimal path of DTW, and the path is smoother and more in line with the actual action similarity calculation.



(a) Modified points after DTW fitting (b) DTW path modified result

Figure. 4 DTW path correction process

The result $D(T,A)$ in the experiment is 317.33, error factor 3000, before introducing the starting and ending point discrimination, the original DTW path is 86.3%, The S (T, A) value of the modified DTW path is 94.7%,After introducing the starting and ending point discrimination, because the modified DTW path starting point decision starts from 5, the behavior similarity

obtained is 79.9%, which is 94.7% compared with the previous 94.7%. The similarity results are more in line with the perception and the volunteers' own feelings.

4. Conclusion

In the algorithm of behavior similarity, the distance between each action information of template behavior and each action information of behavior to be tested is calculated, then the matching path is planned by DTW algorithm, and the DTW path is modified. In the later experiment, the information of different behaviors of 60 groups of different people was randomly selected. The data showed that when the speed of the same behavior was inconsistent, the result of behavior similarity after DTW path optimization was 10.5% lower than that of DTW optimization. For the improvement of DTW slope, compared with the traditional DTW path slope, the slope can only be between 0.5 and 2, which also solves the time between matching and template. A problem that is too bad. Therefore, this method can quickly and easily solve the behavior similarity with different stations, different individuals, excessive behavior time difference, and the stability error of the results is small, which is more in line with the observation results. It is an efficient and simple method to calculate the similarity of human behavior.

Acknowledgement

The authors are grateful to the support of the National Natural Science Foundation of China (No.61373004, No.61501297).

References

[1] Suraj Prakash Sahoo, Samit Ari. On an algorithm for human action recognition[J]. Expert Systems with Applications, 2019, 115: 524-534.

[2] Maity, Satyabrata, Bhattacharjee, et al. A Novel Approach for Human Action Recognition from Silhouette Images[J]. IETE Journal of Research, 2017, 63(2): 160-171.

[3] Beaudry, Cyrille1, Péteri, et al. An efficient and sparse approach for large scale human action recognition in videos[J]. Machine Vision and Applications,2016, 27(4): 529-543.

[4] Mona, Elsayed, et al. An enhanced method for human action recognition[J]. Journal of Advanced Research, 2015, 6(2): 163-169.

[5] Asteriadis, Stylianos, Daras, et al. Landmark-based multimodal human action recognition[J]. Mutimedia Tools ans Applications, 2017, 76(3): 4505-4521.

[6] Xu Zongyi, Zhang Qianni, Cheng Shiyang. Multilevel active registration for kinect human body scans: fromlow quality to high quality[J]. Multimedia Systems, 2018, 24(3): 257-270.

[7] Seddik, Gazzah. Human-action recognition using a multi-layered fusionscheme of Kinect modalities[J]. Iet Computer Vision, 2017, 11(7):530-540.

[8] Liu Xueping, Li Yibo, Li Xiaoming, et al. Multi-feature consultation model for human action recognition in depth video sequence[J]. The Journal of Engineering. 2018, (16): 1498-1502.

[9] Gavrilova, Wang Yingxu, Ahmed, et al. Kinect Sensor Gesture and Activity Recognition: New Applications for Consumer Cognitive Systems [J]. IEEE Consumer Electronics Magazine, 2017, 7(1): 88-94.

[10] Li Qiannan, Wang Yafang, Sharf, et al. Classification of gait anomalies from Kinect[J]. Visual Computer, 2018, 34(2): 229-241.

[11] Zhu Guangming, Zhang Liang, Shen Peiyi, et al. An Online Continuous Human Action Recognition Algorithm Based on the Kinect Sensor[J]. Sensors, 2016, Vol.16 (2), doi:10.3390/s16020161.