

A Gestures Trajectory Recognition Method Based on DTW

Duan Hong*

Software School of Xiamen University
Xiamen city, China

*corresponding author: hduan@xmu.edu.cn

Luo Yang

Software School of Xiamen University
Xiamen city, China

e-mail: 779446731@qq.com

Abstract—This paper puts forward a gestures trajectory recognition method based on DTW. Through the direction characteristic to calculate trajectory characteristics, and through the coding to quantize direction characteristic, and at the same time, considering the direction of the cyclical, proposed a new distance equation for calculating distance. The experimental results prove that the method realized the dynamic gesture recognition in the complex static background.

Keywords—Gestures Trajectory Recognition, DTW, Gestures Features

I. INTRODUCTION

Human-computer interaction technology's development makes the gesture recognition technique become a hot research spot [1]. Among them, bare hand gesture recognition based on visual monocular has natural, direct and concise characteristic relative other man-machine interactive way [5, 6], so it has a broad prospect of application in computer games, robot control and home appliance control. And in these applications, gestures trajectory identification becomes the key to the whole technology application [7-10].

At present the most common gestures trajectory recognition technology is based on the recognition of DTW [1] and the HMM [2]. The first of these methods mainly used in the speech recognition, and has good recognition effect. HMM mainly used in continuous gesture recognition, and it has higher success rate in complex involved in the context of the gesture recognition, but it need a lot and repeated training, according to this paper's dynamic gesture recognition requirements, using DTW algorithm is more simple, easy to realize, and can satisfy the real time requirement [3].

II. DTW ALGORITHM

Dynamic gesture is composed by a series of gestures trajectory. Since the different speed of the different outgoing gesture, we get different number of Characteristic sequence. For example, there is the following two characteristic sequences, $A=\{1,2,3,4,10,3\}$, $B=\{1,2,3,10,4,3\}$. It should be 8.45 if we use Euclidean distance to calculate the total distance. However the sequences of images represented by the two sequences are very similar, since they are different in the timeline. If we change the sequence A to $\{1,1,1,10,2\}$, like the length of A and B is different, we can't use the Euclidean distance to calculate. So in this paper, we use the DTW technology based on dynamic programming technique to solve this problem.

DTW algorithm's main idea is to eliminate the difference between the two different lengths of time, using non-linear regularization function to model the volatility on the timeline, by changing one of the timeline as much as possible with the addition of a timeline of coincidence. During the implement process, the DTW algorithm changes the complex overall optimization problem into many simple local optimization problems [2].

We call the pre-defined the trajectory template as reference template. A reference template can be expressed as:

$$R = \{R(1), R(2), \dots, R(m), \dots, R(M)\},$$

m represents frame label of gestures trajectory. $m=1$ represents the starting frame; $m=M$ represents the ending frame; M represents the total number of frame label in this template. $R(m)$ is the feature amount of the No. m frame. The gesture trajectory to be identified called test template, can be expressed as $T = \{T(1), T(2), \dots, T(n), \dots, T(N)\}$.

n is the label of the test frame. $n=1$ represents the starting frame; $n=N$ represents the ending frame; N represents the total number of frame label in this test template. $T(n)$ is the feature amount of the No. n frame. The reference template generally uses the same type of feature amount and the same frame length with the test template.

In order to compare the similarity between reference template and test template, we need to compute their distance $D[T, R]$. The shorter the distance, the higher the similarity. In this system, we use metrics based on Euclidean distance. For compute the similar distance, we should count from the distance of all the frames in reference template and test template. Suppose n and m is respectively the frame number arbitrarily selected in T and R , $d[T(n), R(m)]$ represents the distance between these two frames.

If N equals M , we can simply calculate the distances between each frame, making the sum finally. But when N is unequal M , we have to consider about aligning the reference template and the test template. The alignment needs linear expansion method, if N is less than M , we will enlarge the test template to a frame sequence lasting M frames first, and calculate the distance between reference template and test template. However, this kind of calculation is not in consideration of the differences in length shown in gesture trajectory under various circumstances, therefore the recognition may not get the best result, which means we are using more dynamic programming.

First, we construct a 2D rectangular coordinate system, marking out the first ~ the M th frames of the reference

template on the Y-axis of the system, the first ~ the Nth frames of the test template on the X-axis of the system. Then we make grids by drawing horizontal and vertical lines through those marked points in the system. Each point(n,m) represents the matching of the Nth frames of the test template and the Mth frames of the reference template, as the figure 1 shows below. The dynamic programming algorithm can be concluded as looking for a path passing through some points in the 2D rectangular coordinate system, where these points means the frames we need take and calculate in the test template and the reference template, besides difference made by this path should be the minimum. In addition, there are restraining conditions for selecting this specific path, first the speed of showing a particular gesture can be changeable but there will still be a sequence in time order, which means the path must starts on the point (1, 1) and ends on the point (n, m).

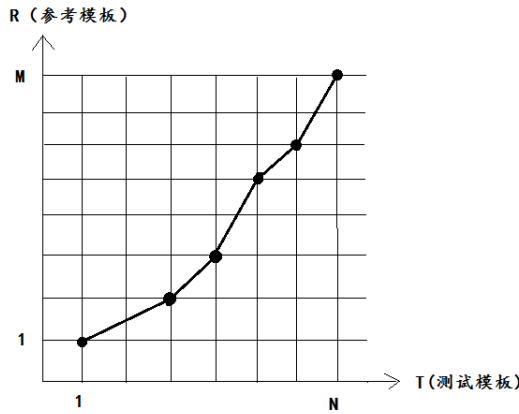


Figure 1. Search path of DTW

Suppose a_i is some frame in the test template, b_j is another frame in the reference template, mark $D(a_i, b_j)$ as the cumulative distance between a_i and b_j . In order to make the path mild and gentle, we can restrain the gradient in the range between 0 and 2, if the path has already passed the point (a_i, b_j) , the former point it has passed should be one of the 3 points in the figure 2 below:

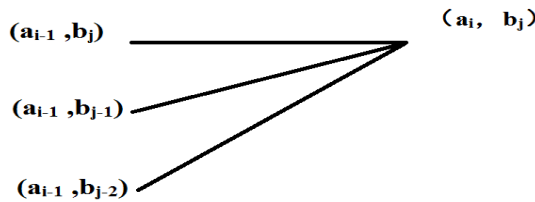


Figure 2. Search path constraint

The recursion equation for calculating the cumulative distance:

$$D(a_i, b_j) = d(a_i, b_j) + \min(D(a_{i-1}, b_j), D(a_{i-1}, b_{j-1}), D(a_{i-1}, b_{j-2})) \quad (1)$$

$D(a_i, b_j)$ means Euclidean distance. In this way, we can start from the point (1, 1), get the best path after repeating recursion.

As for the realization, we allocate $2 \times N \times M$ matrices, for cumulative distance matrix D and Frame-Matching distance matrix d . Among the matrix d , the value of $d(i, j)$ is the Euclidean distance between the characteristic of the first i frame of the test template and the characteristic of the first j frame of the reference template. $D(N, M)$ is the distance of the best matching path.

III. THE DECISION OF THE FEATURE OF TRACK

First we need to mark out the start and end of gesture process. We calculate the distance between the gestures of current second and last second. When the distance is bigger than some certain threshold, it indicates that the gesture-process is started. When the distance is smaller than the threshold, it indicates that the gesture-process is over. And suppose now we have a track which is linked by n points. If the value of n is too small, it can be a situation that is caused by hand-shaking, so it wouldn't be distinguished, or the DTW algorithm will distinguish it.

In fact, the feature of track is the feature of direction [4], so we reorganize the feature to the feature of direction. We defined 16 directions, then number the direction of track, as shown in Figure 3:

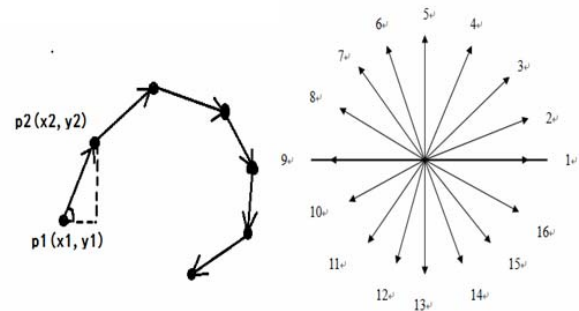


Figure 3. The picture which number the direction

When we calculate distance, if one direction is numbered as 1, the other is 16, the deviation value is 15. In fact, shown in the picture, they are two adjoining directions with one-unit difference, so we use equation (2) to calculate the difference between two directions:

$$d = \begin{cases} |v1 - v2| & \text{if } (|v1 - v2| \leq 8) \\ 16 - |v1 - v2| & \text{if } (|v1 - v2| > 8) \end{cases} \quad (2)$$

In the equation, $v1$ and $v2$ are numbers of two directions, d is the distance we calculated.

IV. RESULT AND CONCLUSION OF THE EXPERIMENT

We define six kinds of gesture process totally, compare the testing track to the template, and calculate the min DTW. The value correspond to the gesture process. At the same time we set a minimum threshold c to avoid the wrong judgment of some undefined gesture. When the value is


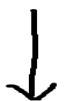



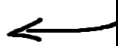
bigger than c , which indicates that the distance between undefined track and model-track is relatively big, and there is no model can be matched, thus we know that is undefined gesture. The graphic 1 is the result of experiments:

We can find out that the accurate rate of our method is basically maintained over 80 percents. And average time of figuring successfully is between 2-3 ms, which can satisfied the instant operating control.

REFERENCES

- [1] K.Takahashi, S.Seki, R.Oka, "Spotting Recognition of Human Gestures from Motion Images," The Inst. of ElectromCS, Information and Comm, vol.36, No.7, pp.28-35, 1993.
- [2] X.D.Huang, Y Aiki, M.A.Jack, "Hidden Markov Models for speech Recognition," Edinburgh UniV. Press, 1990.
- [3] LIU Jiang-hua, Cheng Jun-shi, Chen Jia-pin, "VISON BASED DYNAMIC GESTURE RECOGNITION AND ITS APPLICATION IN HUMAN-HUMANOID ROBOT INTERACTION," vol.24, No.3, pp.197-200, 2002.
- [4] LUO Yang, "Gesture Recognition based on Skin Color and Background Difference," Xiamen University, bachelor's degree thesis, 2012.
- [5] GUO Kang-de, "3D fingertip detection algorithm and application based on visual," Zhejiang University, 2010.
- [6] W Andrew, S Mubarak, D Vitoria, "A virtual 3D blackboard: 3D finger tracking using a single camera," IEEE International Conference on Automatic Face and Gesture Recognition, pp.536-543, 2000.
- [7] LIU Jun-mei, "Vision-based gestures recognition," Beijing Jiaotong University, 2006.
- [8] LU Kai, LI Xiao-jian, "Gesture recognition research overview," Journal of Xi'an University of Arts & Science, vol.9, No.2, pp.91-94, 2006.
- [9] REN Hai-bing, ZHU Yuan-xin, XU Guang, LIN Xue, ZHANG Xiao-ping, "Vision-based recognition of hand Gestures-A survey," Acta Electronica Sinica, No.2, pp.118-121, 2000.
- [10] Mathias K'olsch, Matthew Turk, "Robust Hand Detection," Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'04), 2004.

TABLE I. THE RESULTS OF DYNAMIC GESTURES FIGURING

Tracking	Number of experiment	Number of figuring out successful	Rate of success	Average figuring time
向上 	20	17	85%	2.15
向下 	20	16	80%	2.22
向左 	20	17	85%	2.18
向右 	20	16	80%	2.45
下转右 	20	16	80%	2.60
下转左 	20	15	75%	2.78