Dynamically Weighted DTW for Dynamic Full-Body Gesture Recognition

Hyo-Rim Choia, HyeYeon Cho, TaeYong Kimb

aGraduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, Republic of Korea.
funappear@cau.ac.kr, nuage100@gmail.com, kimen@cau.ac.kr
b*Corresponding Author: kimen@cau.ac.kr

Abstract

As sensor technology has become more advanced, research projects exploring new ways in human-computer interaction have been enabled. Especially, after the Kinect released, study of gesture recognition for human-computer interaction using Kinect has become popular. Dynamic Time Warping (DTW) is one of the template matching algorithm used in the field of gesture recognition. By selecting the most similar gesture compared with the DTW to input sequence and reference sequence of the database, it is possible to achieve a gesture recognition. When trying to recognize a dynamic full-body gesture in Kinect environment, we have to compare all joints throughout the body. When calculating the similarity of the two sequences, all joints of the body is not important in the same way. We propose a dynamically weighted DTW method based on the displacement of the joint. The proposed method shows good performance in the experimental results.

Keywords: Dynamic Time Warping, Gesture Recognition, dynamic Weight, Kinect.

1. Introduction

Generally used human-computer interface are mouse and keyboard. However, recently, natural user interface (NUI) based on user motion is being more applied due to the improvement of the sensor technology. After Microsoft Kinect released, NUI studies utilizing depth camera has become popular. Gesture recognition is an important technology used in NUI. Many methods have been proposed for gesture recognition, including the Dynamic Time Warping (DTW), Hidden Markov Models (HMM) and Conditional Random Fields (CRF). Dynamic time warping algorithm is a pattern matching algorithm that allows a nonlinear stretching of the sequence. To recognize the human gesture, DTW calculating the similarity between input data and predefined reference sequence. In case of recognizing full-body gesture, all body joints are not equally important. In this paper1,2,3, we introduce DTW approaches for full-body gesture and our method.

2. DTW gesture recognition for full-body gesture

2.1 Standard dynamic time warping

DTW algorithm measures the similarity between any two variable sequences1. When we compare sequence R and T (R has N length, T has M length), we can generate an N-by-M matrix g which is the accumulated cost matrix.

\[
g(n,m) = d(n,m) + \min \begin{cases} g(n,m-1) \\ g(n-1,m-1) \\ g(n-1,m) \end{cases} \quad (1)
\]

Typically function \( d \) is the Euclidean distance. Calculated distance value is \( g(N,M) \).

In order to compare two full-body sequences, calculate average of whole DTW distance which is calculated from each joint.

2.2 Locally weighted DTW

Locally Weighted DTW4 consider displacement of each joint.

\[
C_g^j = \sum_{n=2}^{N} \text{Dist}^j(f_{n-1}^g, f_n^g), \quad (2)
\]

where \( g \) is the gesture index, \( j \) is the joint index and \( n \) is the frame number. \( \text{Dist}^j \) computes the displacement of \( j \)th joint's two consecutive coordinates in feature vector \( f_n \).
and \( f_b \).

After the total displacements are calculated, filter out the noise.

\[
C_i^j = \begin{cases} 
C_a & \text{if } 0 \leq C_i^j < T_1 \\
\frac{C_i^j - T_1}{T_2 - T_1} (C_b - C_a) + C_a & \text{if } T_1 \leq C_i^j < T_2 \\
C_b & \text{otherwise,}
\end{cases}
\]

where \( C_a \) and \( C_b \) are threshold values, and \( T_1 \) and \( T_2 \) are experimentally determined boundary values for threshold assignment. The weight of gesture class \( g \) are calculated via

\[
lw_j^g = \frac{1 - e^{-\beta c_i^j}}{\sum_k (1 - e^{-\beta c_i^j})},
\]

Where \( lw_j^g \) is joint \( j \)'s weight value for gesture class \( g \). Weight \( lw_j^g \) is multiplied to the \( d \) in DTW. Parameter \( \beta \) is empirically determined. Measure the \( Dw(\beta) \) which is sum of average DTW distance between gesture classes and \( Dw(\beta) \) which is sum of average DTW distance within-class. And then discriminant ratio of a given \( \beta \), \( R(\beta) \), is then obtained by

\[
R(\beta) = \frac{D_w(\beta)}{D_v(\beta)},
\]

The optimum \( \beta \) is chosen as the one that maximizes \( R \).

2.3 Globally weighted DTW

Globally weighted DTW\(^5\) use each joint’s average DTW distance between gesture classes and within-class. Global weights is given by:

\[
gw_j = \max (0, \frac{D_v^j}{D_w^j}),
\]

where \( D_v \) is average DTW distance between gesture classes corresponding to joint \( j \), \( D_w \) is average DTW distance within-class corresponding to joint \( j \).

2.4 Enhanced globally weighted DTW

Enhanced globally weighted DTW\(^6\) is GDWT’s extension version. EGDTW additionally calculate the weight of each classes. Weight of EDTW is obtained by

\[
gw_j^e = \max (0, \frac{D_v^j - D_w^j}{D_v^j}),
\]

where \( D_v \) is average DTW distance between gesture classes corresponding to joint \( j \), \( D_w \) is average DTW distance within-class corresponding to joint \( j \).

2.5 Dynamically weighted DTW

Locally weighted DTW and globally weighted DTW methods have been proposed because all joints of full-body gesture is not equally important. This method determines the weights to each joint. The value of the weight is the same value from the start to the end of the operation. But, considering the movement from the start to the end of the gesture, it is not equally important from the start to the end of the operation.

Therefore, we propose, a method of weight changes according to time. The dynamic weights of the class \( g \) is defined as follows:

\[
vw_j^g = \max \left\{ \alpha, \frac{\sum Dist^i (f_g^j - f_{g_{n+1}}^j)}{\sum Dist^i (f_g^j - f_{g_{n+1}}^j)} \right\}
\]

Where \( g \) is gesture index, \( j \) is joint index and \( n \) is frame number. \( \alpha \) is minimum contribution. Parameter \( \alpha \) is calculated as Equation (4).

3. Experimental result

Table 1. Gesture recognition result

<table>
<thead>
<tr>
<th>Method</th>
<th>Average true positive rate</th>
<th>True positive rate of Fox gesture</th>
<th>True positive rate of Maps gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDTW</td>
<td>0.631</td>
<td>0.533</td>
<td>0.782</td>
</tr>
<tr>
<td>GDTW</td>
<td>0.789</td>
<td>0.692</td>
<td>0.818</td>
</tr>
<tr>
<td>LDTW</td>
<td>0.766</td>
<td>0.684</td>
<td>0.832</td>
</tr>
<tr>
<td>EGDTW</td>
<td>0.864</td>
<td>0.882</td>
<td>0.923</td>
</tr>
<tr>
<td>VWDTW</td>
<td>0.741</td>
<td>0.582</td>
<td>0.889</td>
</tr>
<tr>
<td>GDTW + VW</td>
<td>0.900</td>
<td>0.703</td>
<td>0.944</td>
</tr>
<tr>
<td>LDTW + VW</td>
<td>0.896</td>
<td>0.707</td>
<td>0.980</td>
</tr>
<tr>
<td>EGDTW + VW</td>
<td>0.943</td>
<td>0.886</td>
<td>0.996</td>
</tr>
</tbody>
</table>
The gesture set used in this experiment is made of 15 kinds of full-body gesture which is motivated from UbiSoft’s Just Dance 2015\(^7\). Used capture device is Kinect. Kinect is operated at 30 frames per second. The Kinect SDK produce body skeleton from which the joint coordinate can be obtained. In this paper, we use x, y, z-coordinates of the joints. Gestures from 8 people are accumulated in the database, each gesture repeated 30 times.

Whole database’s sequences are normalized by average of two shoulder’s position.

In front of the measurement of recognition rate, we select the sequence to represent the class. We choose the sequence which has lowest Within-class distance. Excluding selected to represents sequence, all sequences are set to the test sequence.

Use the 1-nearest neighbor\(^8\) method, it was measured
gesture recognition rate. Table 1. show the recognition rate of the experimental results. When used in combination with other methods and when used alone, Performance showed improvement on average (10.75% improved).

Figure 1 and Figure 2 shows a joint displacement graph and Example gesture of a database. Watch advantages when using a dynamic weight, Advantage in Fox gesture (2.15%) is smaller than in the case of Maps gesture (11.35%). If there is a feature that can be distinguished from other gesture to displacement of every time, it is judged to be able to take a high advantage.

4. Conclusions

We have developed a dynamically weighted DTW method to improve recognition rate of gesture recognition. Experimental result show that dynamic weighting method improve recognition rate. In gestures each joints movement are continued in order, dynamic weighting method, showed a good performance. But, in gestures many joint moves at the same time, relatively performance improvement width was small.

Acknowledgment

This work was supported by the National Research Foundation of Korea grant (MOE-NRF-2014S1A5B6037633) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2013-009166)

References