Consideration about the Application of Dynamic Time Warping to Human Hands Behavior Recognition for Human-Robot Interaction

Ji-Hyeong Han and Jong-Hwan Kim

Department of Electrical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, 305-701, Republic of Korea, {jhhan, johkim}@rit.kaist.ac.kr

Abstract. To prepare the age when humans and robots live together, robots need to understand the meaning of human behaviors for the natural and rational human-robot interaction (HRI). The robot particularly needs to recognize the human hands behavior, since humans usually express their meanings and intentions by using two hands. In this paper, the robot recognizes the human hands behavior by simulating it based on robot's own hands behaviors set and finding the most similar one as human behavior using dynamic time warping (DTW) algorithm. To consider the effects of different variables, i.e. data normalization methods and local cost measures for DTW algorithm, this paper considers two different normalization methods and four different local cost measures and their effects are discussed. The robot successfully recognizes the eight different human hands behaviors by DTW algorithm with the chosen normalization methods and local cost measures.

Keywords: Human hands behavior recognition, dynamic time warping algorithm, human-robot interaction.

1 Introduction

Due to the rapid development of robot technology and intelligence technology [1], [2], humans and robots will live together in no distant future. Humans can have the benefit of living together with robots after solving the most important issue, i.e. the natural and effective human-robot interaction (HRI). Accordingly the research dealing with HRI are exploding and covering various research fields [3]-[13]. For the natural and effective HRI, robots should understand the meaning and intention of human behaviors. Humans use usually both of hands to express the meanings and intentions, therefore robots need to recognize the human hands behavior above all.

In this paper, the robot recognizes the human hands behavior by simulating perceived human behavior data based on its own behaviors. Since it should be able to recognize the human hands behavior independent of the speed, time variation, and length of the behavior, the dynamic time warping (DTW) algorithm is used to find out the most similar behavior, which avoids loss of data by

linear mapping functions [14], [15]. The DTW algorithm was used originally to compare speech patterns for speech recognition [16]. After the first introduction, the DTW algorithm has been applied to many different fields, such as signature recognition, gesture recognition, and data mining [17]-[19]. This paper considers how the DTW algorithm is applied to recognize human hands behaviors by a robot with different data normalization methods and local cost measures. To compare the perceived human hands behavior data with the robot's own behavior data, both data must be normalized since they have different ranges. In this paper, two different normalization methods, i.e. using data range and with zero mean and unit std, are considered and the human hands behavior recognition results using them are discussed. Also, four different local cost measures, which are needed in the DTW algorithm to compare the data sequences, are considered. They are based on 1-norm, 2-norm, infinity norm distances of Minkowski distance and derivative of data, and the effects of different local cost measures on recognition results are discussed.

This paper is organized as follows. Section 2 presents the DTW algorithm and how it is applied to recognize the human hands behavior. In Section 3, experimental results are discussed. Finally, concluding remarks follow in Section 4.

2 Application of DTW Algorithm to Recognize Human Hands Behaviors by a Robot

In this section, the DTW algorithm and its application to recognize human hands behaviors by a robot are explained. The robot perceives the human hands behavior by RGB-D camera sensor and recognizes it by simulating it based on robot's own hands behaviors set and finding the most similar one by DTW algorithm. Since the hands behavior data should be normalized before applying the DTW algorithm, two different normalization methods are considered. Also, the local cost measure for DTW algorithm is needed to define, therefore four different local cost measures are considered.

2.1 DTW Algorithm

The objective of the DTW algorithm is to find an alignment between two time series sequences with a minimal overall cost. Suppose there are two time series sequences $X = \{x_1, x_2, \cdots, x_N\}$ and $Y = \{y_1, y_2, \cdots, y_M\}$. To find an alignment between two different sequences, the local cost measure $cost(x, y)$ is needed to compare the sequences. The local cost measure for each pair of elements of the sequences X and Y is evaluated and then the local cost matrix $CM \in \mathrm{R}^{N \times M}$ that is defined as $CM(n, m) = cost(x_n, y_m)$ is obtained. The goal is finding the alignment between X and Y with a minimal overall cost, for which three conditions in the following definition should be satisfied.

Definition 1. *An (N, M)-warping path is a sequence $p = (p_1, \cdots, p_K)$ with $p_k = (n_l, m_l) \in [1 : N] \times [1 : M]$ for $k \in [1 : K]$ satisfying the following*

conditions.
(i) Boundary condition: $p_1 = (1,1)$ *and* $p_K = (N,M)$.
(ii) Monotonicity condition: $n_1 \leq n_2 \leq \cdots \leq n_K$ *and* $m_1 \leq m_2 \leq \cdots \leq m_K$.
(iii) Step size condition: $p_{k+1} - p_k \in \{(1,0),(0,1),(1,1)\}$ *for* $k \in [1:K-1]$.

2.2 Applying DTW to Recognize Human Hands Behavior

By using the DTW algorithm, the robot simulates the obtained human hands behavior data from a RGB-D camera sensor in its mind to find out the most similar behavior in its own hands behaviors set. The behavior data of both human hands are defined as $H_{L/R} = \{H_{L/R_1}, \cdots, H_{L/R_N}\}$, where $H_{L/R_n} = (H_{L/RX_n}, H_{L/RY_n}, H_{L/RZ_n})$ and $n \in [1:N]$. The behavior data of both robot hands are defined as $R_{L/R} = \{R_{L/R_1}, \cdots, R_{L/R_M}\}$, where $R_{L/R_m} = (R_{L/RX_m}, R_{L/RY_m}, R_{L/RZ_m})$ and $m \in [1:M]$. The human left and right hands behaviors data are compared with robot's left and right ones, respectively, and all data are time series sequences in (x, y, z) coordinates.

Since the ranges of human hands behavior data and robot's hands behavior data are different and they are three dimensional, they should be normalized before comparing. There can be several normalization methods and in this paper two methods are considered. The first one is normalizing the data to $[0,1]$ by using data range. The range of human hands behavior data, HR, is calculated as:

$$HR_X = \max(H_{LX}, H_{RX}) - \min(H_{LX}, H_{RX}) \tag{1}$$
$$HR_Y = \max(H_{LY}, H_{RY}) - \min(H_{LY}, H_{RY}) \tag{2}$$
$$HR_Z = \max(H_{LZ}, H_{RZ}) - \min(H_{LZ}, H_{RZ}) \tag{3}$$
$$HR = \max(HR_X, HR_Y, HR_Z). \tag{4}$$

Then the normalized human hands behavior data, $\overline{H}_{L/RX_n}$, $\overline{H}_{L/RY_n}$, and $\overline{H}_{L/RZ_n}$, are calculated as follows:

$$\overline{H}_{L/RX_n} = \frac{H_{L/RX_n} - \min(H_{LX}, H_{RX})}{HR} \tag{5}$$
$$\overline{H}_{L/RY_n} = \frac{H_{L/RY_n} - \min(H_{LY}, H_{RY})}{HR} \tag{6}$$
$$\overline{H}_{L/RZ_n} = \frac{H_{L/RZ_n} - \min(H_{LZ}, H_{RZ})}{HR}. \tag{7}$$

The other one is normalizing the data as having zero mean and unit standard deviation. The normalized human hands behavior data are calculated as follows:

$$\overline{H}_{L/RX_n} = \frac{H_{L/RX_n} - mean(H_{LX}, H_{RX})}{std(H_{LX}, H_{RX})} \tag{8}$$
$$\overline{H}_{L/RY_n} = \frac{H_{L/RY_n} - mean(H_{LY}, H_{RY})}{std(H_{LY}, H_{RY})} \tag{9}$$
$$\overline{H}_{L/RZ_n} = \frac{H_{L/RZ_n} - mean(H_{LZ}, H_{RZ})}{std(H_{LZ}, H_{RZ})}. \tag{10}$$

The robot's hands behavior data are also normalized in the same way as human ones.

The local cost measure $cost(\overline{H}_{L/R}, \overline{R}_{L/R})$ for the DTW algorithm can be defined as several distance measures. In this paper, three kinds of distance measures are considered, i.e. 1-norm, 2-norm, and infinity norm distances of Minkowski distance. Minkowski distance-based local cost measures are defined as follows in order of 1-norm, 2-norm and infinity norm distances:

$$cost(\overline{H}_{L/R_n}, \overline{R}_{L/R_m}) = \begin{cases} |\overline{H}_{L/RX_n} - \overline{R}_{L/RX_m}| + |\overline{H}_{L/RY_n} - \overline{R}_{L/RY_m}| + |\overline{H}_{L/RZ_n} - \overline{R}_{L/RZ_m}| \\ \sqrt{(\overline{H}_{L/RX_n} - \overline{R}_{L/RX_m})^2 + (\overline{H}_{L/RY_n} - \overline{R}_{L/RY_m})^2 + (\overline{H}_{L/RZ_n} - \overline{R}_{L/RZ_m})^2} \\ \max(|\overline{H}_{L/RX_n} - \overline{R}_{L/RX_m}|, |\overline{H}_{L/RY_n} - \overline{R}_{L/RY_m}|, |\overline{H}_{L/RZ_n} - \overline{R}_{L/RZ_m}|). \end{cases} \tag{11}$$

Also, a derivative measure can be used as the local cost measure [20]. The estimated derivative for human hands behavior data is defined as follows:

$$der(\overline{H}_{L/RX,Y,Z_n}) = \frac{(\overline{H}_{L/RX,Y,Z_n} - \overline{H}_{L/RX,Y,Z_{n-1}}) + (\overline{H}_{L/RX,Y,Z_{n+1}} - \overline{H}_{L/RX,Y,Z_{n-1}})/2}{2}. \tag{12}$$

The estimated derivative for robot's hands behavior data is also calculated in the same way as human ones. Because the estimated derivatives for the first and last elements are not defined in the above formula, they are the same as the estimated derivatives of second and penultimate elements. Then, the derivative-based local cost measure is defined as follows:

$$\begin{aligned} &cost(\overline{H}_{L/R_n}, \overline{R}_{L/R_m}) \\ &= (der(\overline{H}_{L/RX_n}) - der(\overline{R}_{L/RX_m}))^2 + (der(\overline{H}_{L/RY_n}) - der(\overline{R}_{L/RY_m}))^2 \\ &\quad + (der(\overline{H}_{L/RZ_n}) - der(\overline{R}_{L/RZ_m}))^2. \end{aligned} \tag{13}$$

The DTW algorithm evaluates the defined local cost measure for each pair of elements of the sequences $\overline{H}_{L/R}$ and $\overline{R}_{L/R}$ and obtains the local cost matrix $CM_{L/R} \in \mathrm{R}^{N \times M}$ that is defined as $CM_{L/R}(n, m) = cost(\overline{H}_{L/R_n}, \overline{R}_{L/R_m})$.

The total cost, $cost_p(\overline{H}_{L/R}, \overline{R}_{L/R})$, of a warping path p with the local cost measure is defined as

$$cost_p(\overline{H}_{L/R}, \overline{R}_{L/R}) = \sum_{k=1}^{K} cost(\overline{H}_{L/R_k}, \overline{R}_{L/R_k}). \tag{14}$$

The optimal warping path between $\overline{H}_{L/R}$ and $\overline{R}_{L/R}$ is p^* having the minimal total cost among all possible warping paths which satisfy three conditions in Definition 1. The DTW distance, $DTW(\overline{H}_{L/R}, \overline{R}_{L/R})$, is defined as follows:

$$\begin{aligned} &DTW(\overline{H}_{L/R}, \overline{R}_{L/R}) = cost_{p*}(\overline{H}_{L/R}, \overline{R}_{L/R}) \\ &= \min\{cost_p(\overline{H}_{L/R}, \overline{R}_{L/R}) |\ p \ \text{ is } \ (N, M)-\text{warping path}\}. \end{aligned} \tag{15}$$

The final DTW distance is the sum of DTW distances for left and right hands behaviors data, i.e. $FDTW = DTW(\overline{H}_L, \overline{R}_L) + DTW(\overline{H}_R, \overline{R}_R)$. The robot calculates the $FDTW$ between the perceived human hands behavior data and all behaviors in its behaviors set. Then, the robot hands behavior with the minimum $FDTW$ is recognized as the human hands behavior. This method has two main advantages that it does not need to learn any data in advance before the recognition of the human hands behaviors and it can be applied to the robot that has different degrees of freedom (DOFs) from the human DOFs.

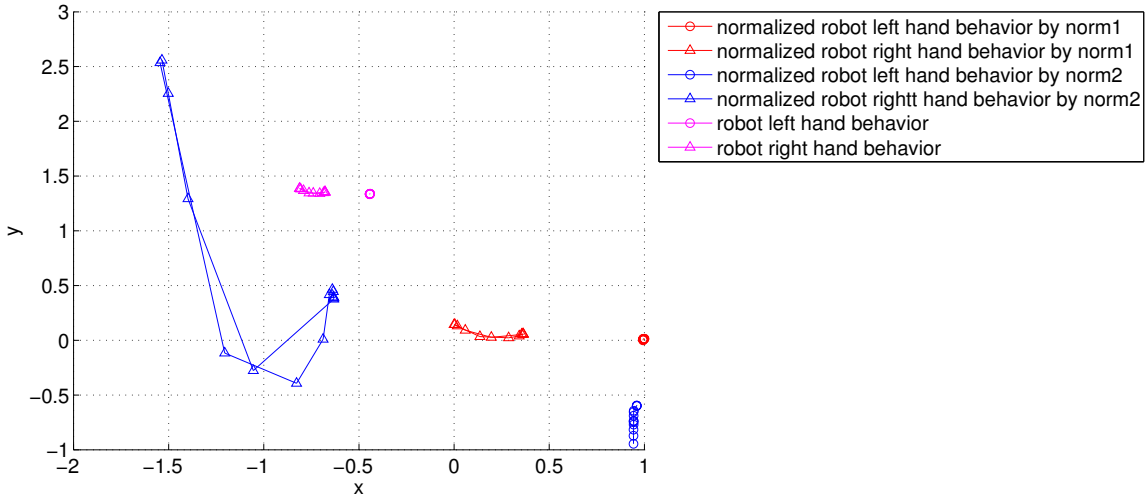
3 Experiments

In this section, experimental setup and results are discussed. The robot's hands behaviors set consisted of eight behaviors, i.e. waving both hands (WB), waving one hand (WO), pointing (Po), touching (Tch), pushing (Pu), grasping (Gr), releasing (Re), and throwing (Th). The robot's hands behavior data were defined as trajectories of robot's hands, which were end-effectors of robot arms, by calculating forward kinematics for each behavior. Since the robot recognized the human hands behavior by simulating it based on robot's own behaviors, the human hands behaviors were recognized as the above eight labels. The basic human hands behavior data set was gathered by a human subject which did the above eight hands behaviors by ten times for each in front of RGB-D camera sensor Kinect. Ten test data for each gathered human behavior were made by adding 0.01 level random noise to each of gathered human hands behavior trajectory data. Therefore, 100 behaviors data for each labeled one and totally 800 behaviors composed the test human behavior data set.

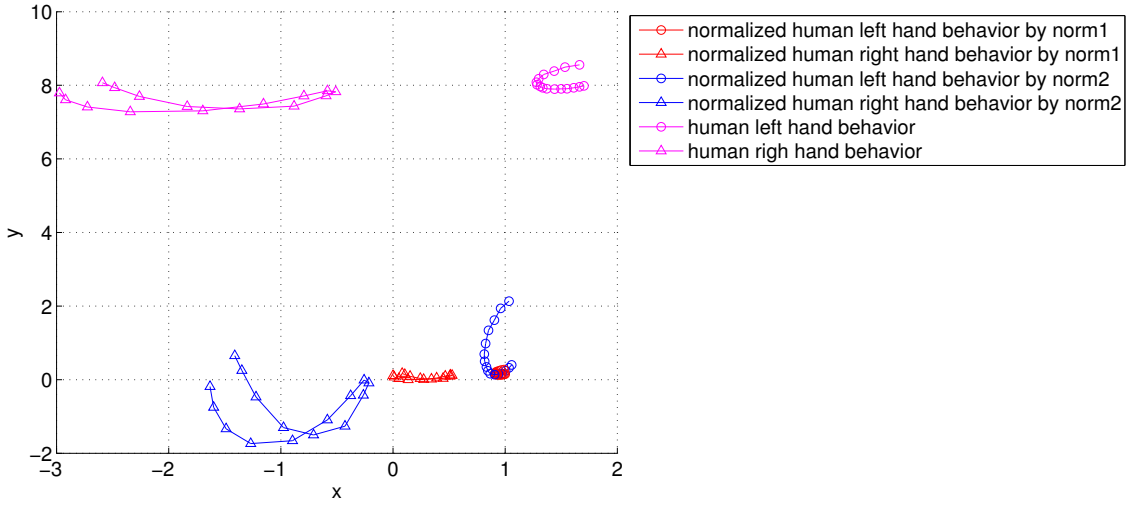
The human hands behavior data was recognized by a robot using DTW algorithm. To consider the effect of different normalization methods and local cost measures to apply DTW algorithm for human hands behavior recognition, the explained two normalization methods and four local cost measures in Section 2 were used and the recognition results were compared.

Table 1. Human hands behavior recognition results using DTW algorithm.

Normalization method	Local cost measure	WB	WO	Po	Tch	Pu	Gr	Re	Th	Average of correct recognition (%)
Normalization using data range (norm1)	1-norm	100	82	80	62	100	100	100	100	90.5
	2-norm	100	89	80	59	100	100	100	100	91
	Infinity norm	100	89	80	59	100	100	100	100	91
	Derivative	69	41	80	79	100	100	0	71	67.5
Normalization with zero mean and unit std (norm2)	1-norm	10	0	30	1	0	58	16	93	26
	2-norm	15	0	30	1	0	58	13	93	26.25
	Infinity norm	15	0	30	1	0	57	11	98	26.5
	Derivative	0	0	30	0	0	100	0	40	21.25



(a)



(b)

Fig. 1. Plot of normalizing WO behavior data by norm1 and norm2 and real data in xy-plane. (a) The robot hands behavior of WO behavior. (b) The human hands behavior of WO behavior.

Table 1 shows the human hands behavior recognition results for combinations of two kinds of normalization methods and four kinds of local cost measures. The

values under each behavior, i.e. WB, WO, Po, Tch, Pu, Gr, Re, Th, means the number of correct recognitions among 100 test data for each one. The difference in 1-norm, 2-norm, and infinity norm distances of Minkowski distance did not affect the performance of human hands behavior recognition in both of DTW algorithm with normalizations using data range (norm1) and with zero mean and unit standard deviation (norm2). The DTW algorithm with derivative-based local cost measure did not show better performance compared to that with Minkowski distance-based one in both of normalization methods. Particularly the method using derivate-based local cost measure did not recognize Re behavior at all, because WB and Re behaviors had the similar derivatives and it classified Re behavior as WB behavior. However, it showed better result in Tch behavior than using Minkowski distance-based local cost measure. Because both of a robot and human subject moved their hands less for Tch behavior than other behaviors, the Minkowski distance based on Euclidean space of Tch behavior did not change much compared to other behaviors. Therefore, the derivative-based local cost measure for DTW algorithm was more effective to recognize the short moving behaviors than Minkowski distance-based local cost measure. The recognition result by DTW algorithm with norm1 showed better performance than that with norm2. Because the normalization method using data range made much more similar pattern with the real data than normalization method with zero mean and unit standard deviation as shown in Fig. 1. Fig. 1 shows the WO behavior data as a representative, since the method using norm2 did not recognize WO at all.

4 Conclusion

In this paper, we considered the application of DTW algorithm to the human hands behavior recognition for HRI. The robot recognized the human hands behavior by simulating the perceived human behavior data based on its own behaviors by the DTW algorithm. By using the DTW algorithm, the robot could recognize the human hands behavior independent of the speed, time variation, and length of behavior. To consider the effect of behavior data normalization method and local cost measure in the DTW algorithm, two kinds of normalization methods and four kinds of local cost measures were considered and tested. The experimental results showed that the DTW algorithm with the normalization method using data range did better behavior recognition than that with zero mean and unit standard deviation. The different norms of Minkowski distance as local cost measure did not affect the recognition performance, but they showed always better overall recognition performances than derivative-based local cost measure. However, the DTW algorithm with derivative-based local cost measure showed better performance than that with Minkowski distance-based ones in short moving behavior like touching. Therefore, the DTW algorithm with data normalization using its range and the local cost measure based on combination of Euclidean space distance and derivative depending on moving

distance of behavior would be better way for a robot to recognize human hands behavior.

Acknowledgement

This research was supported by the MOTIE (The Ministry of Trade, Industry and Energy), Korea, under the Technology Innovation Program supervised by the KEIT (Korea Evaluation Institute of Industrial Technology) (10045252, Development of robot task intelligence technology that can perform task more than 80% in inexperience situation through autonomous knowledge acquisition and adaptational knowledge application).

Also this research was supported by the MOTIE (The Ministry of Trade, Industry and Energy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA (National IT Industry Promotion Agency) (H1502-13-1001, Research Center for Robot Intelligence Technology).

References

1. J.-H. Kim, S.-H Choi, I.-W.Park, and S.A. Zaheer, "Intelligence technology for robots that think," *IEEE Comput. Intell. Mag.*, Aug. 2013, to be published.
2. J.-H. Kim, W.-R. Ko, J.-H. Han, and S.A. Zaheer, "The degree of consideration-based mechanism of thought and its application to artificial creatures for behavior selection," *IEEE Comput. Intell. Mag.*, vol. 7, no. 1, pp. 49-63, Jan. 2012.
3. N. Kubota and K. Nishida, "Perceptual Control Based on Prediction for Natural Communication of a Partner Robot," *IEEE Trans. on Industrial Electronics*, vol. 54, no. 2, pp. 866-877, 2007.
4. E. Sato, T. Yamaguchi, and F. Harashima, "Natural Interface Using Pointing Behavior for Human-Robot Gestural Interaction," *IEEE Trans. on Industrial Electronics*, vol. 54, no. 2, pp. 1105-1112, 2007.
5. C. Mitsantisuk, S. Katsura, and K. Ohishi, "Force Control of Human-Robot Interaction Using Twin Direct-Drive Motor System Based on Modal Space Design," *IEEE Trans. on Industrial Electronics*, vol. 57, no. 4, pp. 1383-1392, 2010.
6. C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Tran. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 3, pp. 569-573, 2011.
7. S. Lallee et al., "Towards a platform-independent cooperative human robot interaction system: III an architecture for learning and executing actions and shared plans," *IEEE Trans. Auton. Mental Develop.*, vol. 4, no. 3, pp. 239-253, 2012.
8. M. Malfaz, A. Castro-Gonzalez, R. Barber, and M.A. Salichs, "A biologically inspired architecture for an autonomous and social robot," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 3, pp. 232-246, 2011.
9. J.-J. Cabibihan, W.-C. So, S. Pramanik, "Human-recognizable robotic gestures," *IEEE Trans. Auton. Mental Develop.*, vol. 4, no. 4, pp. 305-314, 2012.
10. A. Yorita and N. Kubota, "Cognitive development in partner robots for information support to elderly people," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 1, pp. 64-73, 2011.

11. P. Andry, A. Blanchard, and P. Gaussier, "Using the rhythm of nonverbal human-robot interaction as a signal for learning," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 1, pp. 30-42, 2011.
12. J.-H. Han and J.-H. Kim, "Human-robot interaction by reading human intention based on mirror-neuron system," in *Proc. 2010 IEEE ROBIO*, 2010, pp. 561-566.
13. J.-H. Han and J.-H. Kim, "Human intention reading by fuzzy cognitive map: A human-robot cooperative object carrying task," in *RiTA 2012*, pp. 127-135.
14. M. Müller, "Dynamic time warping," *Information Retrieval for Music and Motion*, Ch. 4, Springer, 2007.
15. P. Senin, "Dynamic time warping algorithm review," *University of Hawaii at Manoa*, USA, 2008.
16. T.K. Vintsyuk, "Speech discrimination by dynamic programming," *Kibernetika*, vol. 4, pp. 8188, 1968.
17. C.C. Tappert, C.Y. Suen, and T. Wakahara, "The state of the art in online handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 8, pp. 787-808, 1990.
18. A. Kuzmanic and V. Zanchi, "Hand shape classification using DTW and LCSS as similarity measures for vision-based gesture recognition system," *Int. Conf. EUROCON*, 2007, pp. 264-269.
19. V. Niennattrakul and C.A. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," in *Int. Conf. Multimedia and Ubiquitous Eng.*, 2007, pp. 733-738.
20. E.J. Keogh and M.J. Pazzani, "Derivative dynamic time warping," in *1st SIAM Int. Conf. on Data Mining*, 2001.