




# AR based Self-sports Learning System using Decayed Dynamic Time Warping Algorithm

A. Ikeda<sup>1</sup> , D. H. Hwang<sup>1</sup>  and H. Koike<sup>1</sup> 

<sup>1</sup>Tokyo Institute of Technology

## Abstract

A self-sports learning system that provides users with real-time multimodal feedback about differences between a user's motion and an expert's motion is proposed. We also propose the Decayed Dynamic Time Warping algorithm, which allows the user to change the motion speed dynamically and repeat a target motion without additional operations. The user can thus imitate an expert's motion conveniently and accurately. The proposed system involves training and replay modes. In the training mode, the system provides audio-visual feedback to help the user imitate the expert's motion. The replay mode allows the user to compare their motion to that of the expert. An augmented reality head-mounted display delivers feedback and provides an immersive three-dimensional training experience.

## CCS Concepts

• **Human-centered computing** → Mixed / augmented reality; Information visualization;



Figure 1: Overview of the proposed system

## 1. Introduction

In several sports, sports-related skills can be improved by imitating an expert's motions. For example, a coach can guide a learner based on their observations of the learner's movements. Learners can also analyze their form by watching videos of a practice session individually or with a coach. For elite athletes, optical or inertial sensor-type motion trackers are used to record motion as three-dimensional (3D) sequential data. The recorded data are visualized using computer graphics, such as a skeleton or full-body model. In addition, advanced image processing software, such as

Dartfish [Dar99], can visually compare two motions by overlapping poses.

With such conventional sports learning systems, however, the speed of the motions may differ in whole or in part. For example, consider a golf swing where the swing speed of the model movement is not always the same as the learner's swing speed. Even if the start and end points of the model movement are adjusted to match those of the learner, differences can persist. Therefore, even if the learner achieves the desired swing trajectory, the learner's movement is still not correct. In addition, providing a method to match user and expert motions in real time is difficult. In the conventional method, the learner watches a video after completing their practice, which helps them recognize faults in their movements. However, as the interval between practice and analysis increases, it becomes increasingly difficult to reproduce previous motions.

Thus, we propose a sport self-learning system that helps learners compare images of their motions to those of an expert in real time regardless of their own movement speed. Figure 1 shows the overview of our proposed system, and our primary contributions are summarized as follows.

- We propose a method based on dynamic time warping (DTW) to match two motion models in real time.
- We implement a learning environment based on augmented reality (AR) that allows the user to observe differences between two motion models from a third-person perspective.

## 2. Related Work

### 2.1. Training systems with real-time feedback

Real-time feedback can enable the user to adjust their motions, poses, and behaviors immediately during training. For this reason, learning systems with real-time feedback have often been proposed.

Hoang et al. proposed Onebody [HRVT16], which displays a skeleton of the user/expert in a first-person perspective using a virtual reality head mounted display (HMD). With Onebody, the user obtains visual feedback by observing differences between their motions and those of an expert. However, with this system, the user and expert must capture the motion simultaneously. Therefore, self-learning is difficult with this system.

The MotionMA [VBG13] system analyzes an expert's motion and converts it to a model, which is then compared to the user's motion. Then, the system provides the user visual feedback. Therefore, the user can learn the correct motion independently. However, the MotionMA system uses a two-dimensional display, such as a television or projector; therefore, feedback is limited relative to directionality (e.g., up, down, left, and right).

### 2.2. Training systems with out-of-body scene

Projecting self-images onto out-of-body scenes and watching it from an external viewpoint are effective to improve sports-related skills [RSM09]. Therefore, several systems employ 3D out-of-body models in AR and Mixed Reality(MR) environments. For example, OutsideMe [YDG\*15] displays a user avatar in an MR environment and enables the user to practice with a virtual expert dancer. Here, the user can observe their entire body from a third-person perspective; consequently, they do not have to turn their head to observe their body.

In My Tai-Chi Coaches [HCZ\*17], the user is surrounded by virtual mirrors that project user and expert avatars into an MR environment. The user can observe the expert's pose projected from various perspectives by simply turning their head. However, this system uses a drone to capture user poses from various angles; therefore, the system configuration is complicated.

Body Cursor [HR17] proposed by Hamanashi displays 3D skeletal models of the user and expert into an AR environment and provides tactile feedback when the user's motion differs from that of the expert. However, the Body Cursor system does not match the user and expert's motions, and the user cannot practice the motion at their own pace.

## 3. Proposed System

Figure 2 shows configuration of the proposed system. An inertial measurement unit-based motion tracking suit (Xsens MVN) is used to capture the motions of the user and an expert, and a HoloLens AR HMD is used to run a motion matching algorithm and provide audio-visual feedback to the user. Note that a personal computer is used to receive the captured motion data and send it to the AR HMD.

As mentioned in Section 1, it is important for a learner to imitate

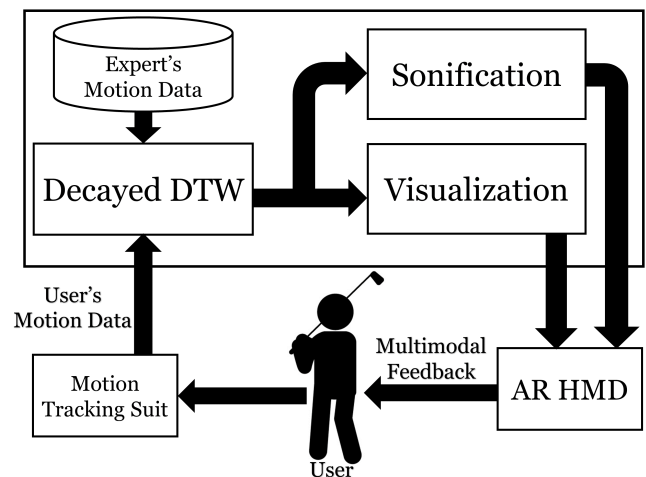


Figure 2: System configuration

an expert's motion at their own pace. To realize this, it is necessary to estimate the position at which the user's current motion matches the expert's entire motion sequence. However, existing systems do not support this feature or support it in a limited manner. Thus, we propose the Decayed DTW algorithm, which accurately estimates the position of the user's motion relative to the entire expert's motion sequence. The proposed algorithm can be run in real time on the standalone AR HMD, which has low computing power. The proposed algorithm is described in the following.

### 3.1. Proposed Decayed DTW algorithm for motion matching

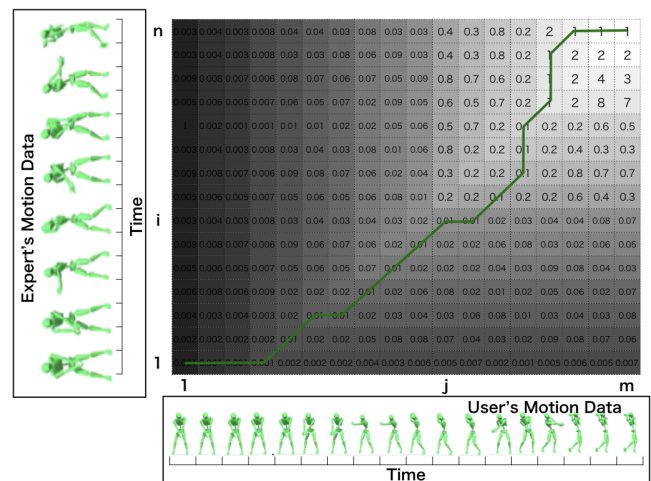


Figure 3: Cost matrix with minimum-cost warp path of Decayed DTW algorithm (the previous cost value is decayed and the influence is decreased).

The system must compare the user and expert's motions to find the current position of the user's motion relative to the entire motion sequence of the expert and visualize differences between the

expert and user's motions. In addition, to imitate the expert's motion correctly, the user performs the motion at their own pace; thus, a robust matching algorithm is required because the motion speed is dynamic.

DTW [BC94] has been used previously to measure the similarity between two temporal sequences. However, when matching data that are updated in real time rather than data of fixed size, the accuracy of the algorithm is reduced; therefore, DTW cannot be utilized in a real-time learning system, such as the proposed system. Thus, to address this problem, we propose Decayed DTW, which decays past matching results according to the following equation.

$$D(i, j) = \min \begin{cases} D(i-1, j) * \beta + d(i, j), \\ D(i-1, j-1) * \beta + d(i, j), \\ D(i, j-1) * \beta + d(i, j) \end{cases} \quad (1)$$

$$0 \leq i < N$$

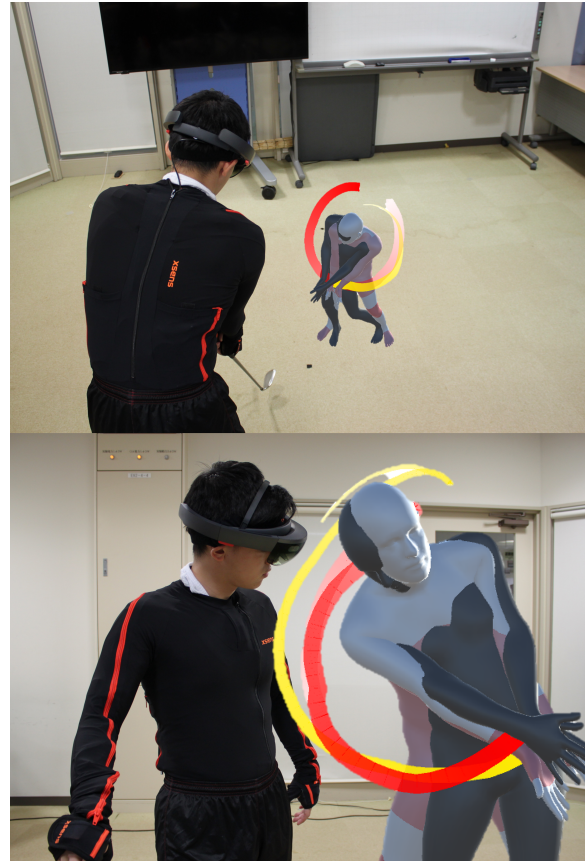
Here,  $i$  is the specific position of the entire expert's motion and  $j$  is the current position of the user's motion.  $N$  is the length of the expert's motion.  $D$  corresponds to the minimum matching cost at position  $(i, j)$ .  $d$  is the error in motion difference calculated using the sum of the Euclidean distances of each joint and  $\beta$  is the decay factor.

As shown in figure 3, a minimum cost is calculated for each frame, and a cost matrix is updated. A position with the minimum matching cost in a current column of the cost matrix is selected as the position at which the user's current motion matches the expert's entire motion sequence and, a current element of minimum-cost warp path. Note that, by decaying past values, matching is not easily affected by noisy motions. The proposed algorithm calculates the current matching cost using the previous calculation result as Eq. (1). The initial value of  $D(0,0)$  is set to  $d(0,0)$ , otherwise,  $D$  is the minimum of values calculated for each expression. The computational and space complexities of each frame are  $O(N)$ .

### 3.2. Audio-visual feedback

The proposed system provides two visualization modes (i.e., training and replay modes) and the user can freely switch between modes. Note that out-of-body models of the user and expert are displayed on the ground of the real world in both modes. The models used in this system are synthesized by SMPL [LMR\*15] and are overlapped and displayed on each other so that the user can recognize differences in the motions. In addition, the degree of difference in a given region between two motions is expressed by the transparency of the color, and the user receives visual feedback to recognize the position and degree of the difference in the motion. In addition, the trajectories of specific body parts are displayed and the transparency is increased over time. The user can thus easily understand changes in motion over time. Figure 4 shows the visualization results of both modes.

In the training mode, the user can quickly recognize differences between motions in real time through a small out-of-body model and audio feedback. The AR HMD used in the proposed system has a limited field of view ( $30^\circ \times 17.5^\circ$ ), and the user cannot perceive



**Figure 4:** Visualization results of the proposed system. The system has training (top) and replay (bottom) modes, and each mode implements feedback methods suited to each training situation.

a lot of visual information during motion. Therefore, the size of the out-of-body model should be sufficiently small to be visible to the user's eyes; thus, a 1/4 scale model, that the entire model takes fits into the field of view, even if the model takes various poses, is used in the proposed system. Since the model is small, the user cannot perceive detailed differences in the motions. Therefore, audio feedback is provided to the user to resolve this problem. Here, the degree of motion error is expressed by the pitch of the sound. Through the audio feedback, the user can recognize the correctness of their motion in real time.

In the replay mode, a motion performed in the training mode can be reviewed through 1:1 scale 3D models, and the differences in the user and expert motions can be perceived in detail. The user can observe the models from free perspective and can easily perceive specific body parts where the motion is incorrect. With these two modes with multimodal feedback, the user can imitate the expert's motion easily and efficiently.

## 4. Pilot Study

A pilot study was conducted to evaluate the proposed system with six university students as subjects. Note that one golf expert was in-



cluded in the subject group. The subjects experienced the conventional method (watching an expert golf swing motion in the video and then following it) and the proposed system. To compare the conventional method and proposed system, we asked subjects the following questions:

- “How did you feel about the usability of the proposed system compared to the common method based on video recording?”
- “In the proposed system, could you easily perceive differences in motion between you and the expert?”
- “Do you have any comments about our system?”

For the first question, all subjects responded positively that the proposed system shows both the user and expert’s motions (the conventional method only shows the expert’s motion). Other subjects responded positively to the fact that their motion was matched to the expert’s motion in real time even when they performed a specific motion repeatedly, and the golf expert in the subject group told us that it is an efficient way to train swing form.

For the second question, some subjects responded that the difference in motion projected on the small model in the training mode was difficult to perceive but the audio feedback compensated the lack of information. Rather than simply overlapping the two models, it was appreciated that incorrect parts of the motion were expressed by the transparency of colors.

Some subjects indicated that wearing a motion tracking suit is uncomfortable. In addition, some subjects criticized the narrow field of view of the AR HMD. Apart from the evaluations, most of the subjects indicated that the system is interesting and easy to engage with.

Based on this feedback, we confirm that, with the proposed system, users can easily perceive differences between the user and expert’s motions, and that the proposed system is more effective relative to learning an expert’s motion compared to the conventional common method.

## 5. Discussion and Future Work

We have proposed an efficient sports learning system to support users in the imitation of an expert’s; however, the proposed system can also be utilized to train various other motions, such as dance, physical rehabilitation, and vocational training. The proposed algorithm is robust relative to repeated motion sequences at any speed and tempo, the proposed system enables the user to repeat certain motions and imitate the expert’s motion easily.

However, we found that some inconvenient elements of the system, such as wearing the motion tracking suit, require improvement. In the pilot study, the subjects complained about that wearing the motion tracking suit and the weight of the AR HMD. Therefore, in the future, we must implement more user-friendly display and motion capture methods. In addition, an objective performance evaluation of the proposed system was not been performed in this study. Thus, we must also conduct experiments to objectively measure the learning outcomes of the proposed system.

## 6. Conclusions

In this paper, we have proposed an immersive self-sports learning system and the Decayed DTW algorithm. The proposed system enables users to imitate an expert’s motion through multimodal feedback, and the user can easily perceive differences in the motions in three dimensions through training and replay modes. Unlike existing systems, even if the user repeatedly performs only a specific motion, the proposed system can accurately match it to the expert’s motion using the Decayed DTW algorithm. The Decayed DTW algorithm can match the user’s motion sequence to the expert’s sequence fast and accurately in real time. The computational load of the proposed algorithm is low because it is based on a dynamic programming approach and uses the previous calculation results in the current calculation. Therefore, it can be used in a standalone AR HMD with low computing power. In a pilot study, we confirmed user convenience and the usability of the proposed system as a motion training system.

## 7. Acknowledgements

This work was supported by JST CREST Grant Number JP-MJCR17A3, Japan.

## References

- [BC94] BERNDT D. J., CLIFFORD J.: Using dynamic time warping to find patterns in time series. In *KDD workshop* (1994), vol. 10, Seattle, WA, pp. 359–370. 3
- [Dar99] DARTFISH: Dartfish, 1999. URL: <https://www.dartfish.com/>. 1
- [HCZ\*17] HAN P.-H., CHEN Y.-S., ZHONG Y., WANG H.-L., HUNG Y.-P.: My tai-chi coaches: An augmented-learning tool for practicing tai-chi chuan. In *Proceedings of the 8th Augmented Human International Conference* (New York, NY, USA, 2017), AH ’17, ACM, pp. 25:1–25:4. doi:10.1145/3041164.3041194. 2
- [HR17] HAMANISHI N., REKIMOTO J.: Body cursor: Supporting sports training with the out-of-body sense. In *CHI WS ’17* (2017), ACM. 2
- [HRVT16] HOANG T. N., REINOSO M., VETERE F., TANIN E.: One-body: Remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (New York, NY, USA, 2016), NordiCHI ’16, ACM, pp. 25:1–25:10. doi:10.1145/2971485.2971521. 2
- [LMR\*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16. 3
- [RSM09] RYMAL A. M., STE-MARIE D. M.: Does self-modeling affect imagery ability or vividness? *Journal of Imagery Research in Sport and Physical Activity* 4, 1 (2009). 2
- [VBG13] VELLOSO E., BULLING A., GELLERSEN H.: Motionma: Motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI ’13, ACM, pp. 1309–1318. doi:10.1145/2470654.2466171. 2
- [YDG\*15] YAN S., DING G., GUAN Z., SUN N., LI H., ZHANG L.: Outsideme: Augmenting dancer’s external self-image by using a mixed reality system. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI EA ’15, ACM, pp. 965–970. doi:10.1145/2702613.2732759. 2