



# An Evaluation Method of Taekwondo Poomsae Performance

Thi Thuy Hoang<sup>1</sup> and Heejune Ahn\*<sup>2</sup>

Department of Electrical and Information Engineering, SeoulTech, Seoul 01811, Republic of Korea

## Abstract

In this study, we formulated a method that evaluates Taekwondo Poomsae performance using a series of choreographed training movements. Despite recent achievements in 3D human pose estimation (HPE) performance, the analysis of human actions remains challenging. In particular, Taekwondo Poomsae action analysis is challenging owing to the absence of time synchronization data and necessity to compare postures, rather than directly relying on joint locations owing to differences in human shapes. To address these challenges, we first decomposed human joint representation into joint rotation (posture) and limb length (body shape), then synchronized a comparison between test and reference pose sequences using DTW (dynamic time warping), and finally compared pose angles for each joint. Experimental results demonstrate that our method successfully synchronizes test action sequences with the reference sequence and reflects a considerable gap in performance between practitioners and professionals. Thus, our method can detect incorrect poses and help practitioners improve accuracy, balance, and speed of movement.

**Index Terms:** Taekwondo Poomsae, human action analysis, 3D computer vision, 3D HPE (human pose estimation), DTW (Dynamic time warping)

## I. INTRODUCTION

Human action analysis is essential in fields including medical treatment, sports, security, and human behavioral science [1]. In this study, we developed a computer-vision-based method for evaluating the Taekwondo Poomsae performance of practitioners using human action analysis. Taekwondo is a traditional Korean martial art, and Taekwondo Poomsae is a series of choreographed movements that simulate various combat situations against imaginary opponents. As a crucial aspect of Taekwondo training, Poomsae is usually judged according to factors such as accuracy, power, speed, and fluidity of movements.

With advances in IT technology, such as computer vision and machine learning, human action evaluation methods have become studied in applications such as dance and con-

ducting [2,3,11]. However, the evaluation of Taekwondo Poomsae poses unique challenges: there is no reference timing information, and poses must be compared between distinct people.

The first important step in human action analysis is to accurately estimate human poses in 2D or 3D space. Many studies have been conducted on human pose estimation (HPE) methods [1], including monocular-[4,5,6] and multi-camera-based [7,8,9] methods. In the present study, we used TCMR [6] because multi-view video data is very scarce for Taekwondo Poomsae. However, because our pose sequence evaluation method is independent of HPE, it yields better performance when used in conjunction with highly accurate 3D HPE methods.

Because human poses are often represented by lists of joints, it is natural to compare poses by directly comparing

Received 27 April 2023, Revised 18 September 2023, Accepted 26 September 2023

\*Corresponding Author Heejune Ahn (E-mail: [heejune@seoultech.ac.kr](mailto:heejune@seoultech.ac.kr))

Department of Electrical and Information Engineering, SeoulTech, 01811, Republic of Korea

Open Access <https://doi.org/10.56977/jicce.2023.21.4.337>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

corresponding joint positions using metrics such as the mean per-joint position error (MPJPE), Procrustes-aligned MPJPE (PA-MPJPE), and percentage of correct keypoints (PCK). However, these metrics cannot be used to compare the poses of different people, owing to variations in limb lengths. Instead, it is better to compare persons with different shapes solely using joint rotations. In [2] the human torso was represented as a 22-dimensional feature composed of a six-dimensional torso feature, an eight-dimensional first-degree feature, and an eight-dimensional second-degree feature. In the present study, inspired by rigged models and the BioVision Hierarchy format [10], we decomposed a human joint representation into pose vectors (joint rotations) and limb offset/length vectors (body shape). Because the resulting pose vectors are shape-independent, they can be used to compare the poses of differently shaped persons.

Relying on this pose vector difference measure, we synchronized the test (practitioner) pose sequence to the reference (professional) pose sequence using dynamic time warping (DTW). Our approach involves synchronizing test action sequences with reference sequences, allowing us to identify performance gaps between professionals and practitioners. Furthermore, our method can identify incorrectly posed joints and action speeds, and can be deployed alongside accurate 3D HPE methods to enhance the balance, coordination, and overall performance of practitioners. We used 3D motion-captured action sequences of the Poomsae world champion [12] as a reference to obtain 3D pose sequences of professionals and practitioners using state-of-the-art monocular 3D human pose estimation and temporal consistent model recovery (TCMR) [6].

The remainder of the paper is organized as follows. Section 2 presents the decomposition method of joints into poses and limbs, the comparison metric, and the DTW-based synchronization method. We present the experimental results using Taekwondo pose sequences in Section 3, and conclude the paper in Section 4.

## II. SYSTEM MODEL AND METHODS

### A. Decomposition of Human Pose Joints

The joint representation of a 3D human pose is defined as a list of 3D positions of predefined joints:

$$\text{Joints: } \{J_j \in \mathbb{R}^3 : j=0, \dots, N-1\}, \tag{1}$$

where  $N$  is the number of joints and  $j$  is the index for each joint. The human joint hierarchy tree is a way to organize the joints in the human body to determine the body's overall posture and movement. The pelvis is typically designated as the root joint, and the connected joints are defined as three

structures. The hips and spine are child nodes of the pelvis, followed by the shoulders, elbows, knees, ankles, and wrists. Finally, there are the “end-effector” joints, which include the hands and feet. Expressing our method in Python-style syntax, the joint tree is defined as a 1D list denoted as  $P$ , where the value  $P[j]$  is the parent joint index  $j$ .

The proposed decomposition method of the human joint representation into joint rotations (posture) and limb lengths (body shape) is defined as the function  $F$ :

$$F: \{J_j \in \mathbb{R}^3 : j=0, \dots, N-1\} \rightarrow \{Pose_j \in \mathbb{R}^3 : j=0, \dots, N-1\} \times \{L_j \in \mathbb{R} : j=0, \dots, N-1\}, \tag{2}$$

using limb lengths, whereas the BVH format uses 3D offsets for the rest of the poses of each joint. The default resting limb position can be represented as  $(0, 0, 0)$ , that is, an offset point along the local z-axis in the BVH style. This decomposition is computed using Algorithm 1. The rotation – that is, the pose for joint  $j$  around joint parent <sub>$j$</sub>  – is defined from three joint positions:  $J[j]$ ,  $J[\text{parent}_j]$ , and  $J[\text{grandparent}_j]$ . However, this rotation is dependent on the parent joints, making it difficult to compare poses between the joint sets of two persons. To address this problem, we defined the local pose. The local rotation or pose of joint  $j$  is defined in the local coordinate frame of its parent joint parent <sub>$j$</sub> .

**Algorithm 1: Decomposition**

*Input: joints position  $J(N, 3)$ , parents list  $P$*

*Output: local poses  $Pose$ , limbs length  $L$*

$Pose[0, :] = J[0, :]$

$rotmat = cv2.Rodrigues(Pose[0, :])[0]$

for  $j$  in range(1, N):

$p\_j = P[j]$

$g\_j = P[p\_j]$  if  $p\_j \geq 0$  else -1

$cur\_vec = J[j] - J[p\_j]$

$prev\_vec = J[p\_j] - J[g\_j]$  if  $g\_j \geq 0$  else  $J[0, :]$

$L[j, 2] = np.linalg.norm(cur\_vec)$

$angle = \arccos((cur\_vec * prev\_vec) / (np.linalg.norm(cur\_vec) * np.linalg.norm(prev\_vec)))$

$axis = (cur\_vec \times prev\_vec) / np.linalg.norm(cur\_vec \times prev\_vec)$

$axis\_angle = angle * axis$

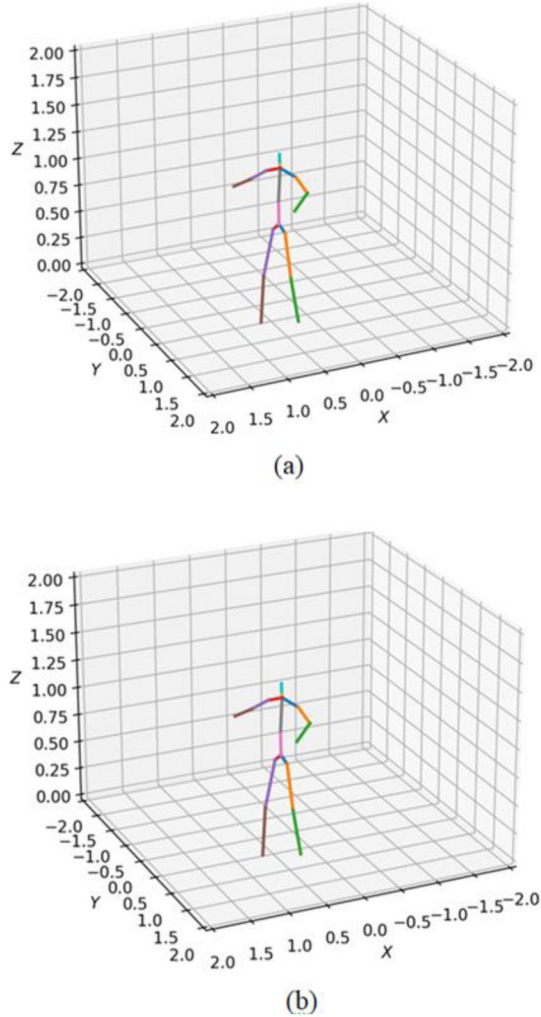
$Pose[j, :] = rotmat * cv2.Rodrigues(axis\_angle)[0]$

We also formulated a pose-to-joint reconstruction method  $F^{-1}$  to reconstruct the original joints, as shown in Algorithm 2. We did not use this reconstruction method to evaluate the Taekwondo Poomsae pose sequences.

$$F^{-1}: \{Pose_j \in \mathbb{R}^3 : j=0, \dots, N-1\} \rightarrow \{L_j \in \mathbb{R} : j=0, \dots, N-1\} \times \{J_j \in \mathbb{R}^3 : j=0, \dots, N-1\}, \tag{3}$$

**Algorithm 2: Reconstruction**  
 Input: local poses  $Pose(N, 3)$ , limbs length  $L$ , parents list  $P$   
 Output: joints position  $J$   
 $J[0, :] = Pose[0, :]$   
 $rotmat[0] = I(3)$   
 for  $j$  in range(1,  $N$ )  
      $p\_j = P[j]$   
      $rotmat[j] = rotmat[p\_j - 1] * cv2.Rodrigues(Pose[j])[0]$   
      $J[j] = J[p\_j - 1] + rotmat[j] * L[j]$

Fig. 1 verifies the results of our method, showing that the reconstructed joints are identical to the original joints, and the same pose list is obtained for a differently shaped person in the same posture. Table 1 lists quantitative results corresponding to Fig. 1.



**Fig. 1.** An example verifying the identity of reconstructed joints corresponding to the same pose between two differently shaped persons: (a) original joints, (b) reconstructed joints from poses and limb lengths.

**B. Pose Comparison Measure**

Because we decompose the joint information to extract the posture, we can compare the poses of two persons irrespective of their shapes. We found that the distance between Rodrigues vectors – that is, axis-angle vectors – is an effective metric for human pose comparison. To compare all joint rotations, we defined the following measure:

$$L_p = \sum_{j=1}^{N-1} w_j \|v_{1,j} - v_{2,j}\| \tag{4}$$

where  $v_{1,j}$  and  $v_{2,j}$  are axis-angle vectors at joint  $j$  of two persons, and  $w_j$  ( $\sum_{j=1}^{N-1} w_j = 1$ ) is the weight factor for joint  $j$ . For simplicity, we used equal weights in our evaluation; i.e.,  $w_j = 1/(N-1)$ .

**Table 1.** Joint, pose, and limbs length values in Fig. 1

Joint ID	Joint name	Joint locations (meter)	Local poses (radian)	Limb lengths (meter)
0	Pelvis	[0.6 0.9 1.0]	[0.0 0.0 0.0]	-
1	Hip (L)	[0.5 1.0 0.9]	[-1.9 -0.9 0.0]	0.1
2	Knee (L)	[0.4 1.0 0.5]	[-0.8 -0.6 0.0]	0.4
3	Ankle (L)	[0.3 1.0 0.1]	[0.0 -0.1 0.0]	0.4
4	Hip (R)	[0.6 0.8 0.9]	[2.0 0.8 0.0]	0.1
5	Knee (R)	[0.7 0.8 0.5]	[0.7 0.7 0.0]	0.4
6	Ankle (R)	[0.8 0.8 0.1]	[0.2 -0.2 -0.0]	0.4
7	Spine	[0.6 0.9 1.2]	[0.0 0.0 0.0]	0.2
8	Neck	[0.5 0.9 1.5]	[0.0 -0.1 0.0]	0.3
9	Neck1	[0.6 0.9 1.5]	[-0.3 0.3 -0.0]	0.1
10	Head	[0.6 0.9 1.6]	[0.1 -0.1 0.0]	0.1
11	Shoulder (L)	[0.4 1. 1.4]	[-1.2 -1.4 -0.0]	0.2
12	Elbow (L)	[0.2 1. 1.2]	[ 0.2 -0.9 -0.0]	0.2
13	Wrist (L)	[0.4 1.1 1.1]	[-1.8 -0.5 -0.0]	0.3
14	Shoulder (R)	[0.7 0.7 1.4]	[1.6 0.9 0.0]	0.3
15	Elbow (R)	[0.9 0.7 1.4]	[-0.4 1.1 -0.0]	0.2
16	Wrist (R)	[1.1 0.8 1.4]	[-0.1 0.1 -0.0]	0.3

**C. Pose Sequence Matching Method**

DTW is a widely used algorithm that compares two time sequences at varying times and speeds. It calculates the optimal alignment between two pose series by stretching or compressing one of the series along the time axis to minimize the distance between corresponding points in the two series. The resulting alignment is used to measure the degree of similarity between the two series. DTW has been applied in a wide range of fields including speech recognition, handwriting recognition, music analysis, and bioinformatics [4]. In this study, we deployed the DTW algorithm to synchro-

nize a test pose sequence with a reference pose sequence. We use  $r(t)$  to denote the frame index of the pose corresponding to the test pose in frame  $t$ .

To implement DTW, we used Equation (4) to measure the distance between test and reference poses as the cost function described in Algorithm 3, with lower values indicating higher accuracy. The resulting matched poses were then analyzed to evaluate the effectiveness of the algorithm and assess differences between each pair of joints in the matching path.

```

Algorithm 3: DTW Pose sequence matching
Input: Target's poses  $p_t(i)$ ,  $i=[1, n]$ , reference's poses  $p_r(j)$ ,  $j=[1, m]$ 
Output: Mean distance, matching path
#Distance matrix
for  $i$  in range(1, n) :
    for  $j$  in range(1, m):
         $costmat[i-1, j-1]= Lp(p_t(i), p_r(j))$ 
#Dynamic time warping
for  $i$  in range(1, n):
     $DTW[i, 0] = inf$ 
for  $j$  in range(1, m):
     $DTW[0, j] = inf$ 
 $DTW[0, 0] = 0$ 
for  $i$  in range(1, n):
    for  $j$  in range(1, m):
         $DTW[i, j] = costmat[i-1, j-1] + \min(DTW[i-1, j],$ 
 $DTW[i, j-1], DTW[i-1, j-1])$ 
#Matching path
 $t = n - 1$ 
 $r = m - 1$ 
 $path = [(t, r)]$ 
while  $t \geq 0$  and  $r \geq 0$ :
     $min\_at = \min(DTW[t-1, r], DTW[t, r-1], DTW[t-1, r-1]);$ 
    if  $min\_at = 0$ 
         $t -= 1;$ 
    if  $min\_at = 1$ 
         $r -= 1;$ 
    if  $min\_at = 0$ 
         $r -= 1, t -= 1;$ 
    if  $t < 0$  or  $r < 0$ :
        break
    else:
         $path.insert(0, (t, r));$ 
#Mean distance
for  $t, r$  in path:
     $distances.append(costmat[t, r])$ 
 $meandist = average(distances)$ 
    
```

### III. RESULTS

#### A. Taekwondo Dataset

There are eight Taekwondo Poomsae forms, each of which represents a different attack or defense technique. In this study, we selected pose accuracy for computerized evaluation.

For the reference pose sequence – that is, the ground truth – we used the 3D Poomsae dataset [12] released in 2020 by the Korea Cultural Information Service Agency (KCISA) (see Fig. 2 for examples). Three-dimensional content for all eight poses was generated by capturing the performance of Taekwondo Poomsae competition winners from around the world using motion capture technology, and then constructing movements through modeling and rigging processes. We obtained the joint list for each frame using the Blender software.

For testing purposes, we used YouTube videos of taekwondo professionals and practitioners performing Poomsae (Fig. 3). To extract 3D joint positions from the videos, we used the TCMR monocular 3D HPE [6], a state-of-the-art algorithm that uses SMPL parameter estimation in a video sequence with a combined CNN-RNN structure. The reported accuracy of TCMR is very high in terms of PAMPIPE, with 41.1 mm for Human3.6m, 55.8 3DPW, and 62.8 mm for MPI-INF-3DHP. However, we found that TCMR cannot correctly estimate certain unusual poses in



Fig. 2. KCISA Reference Sample Data [KCISA20]

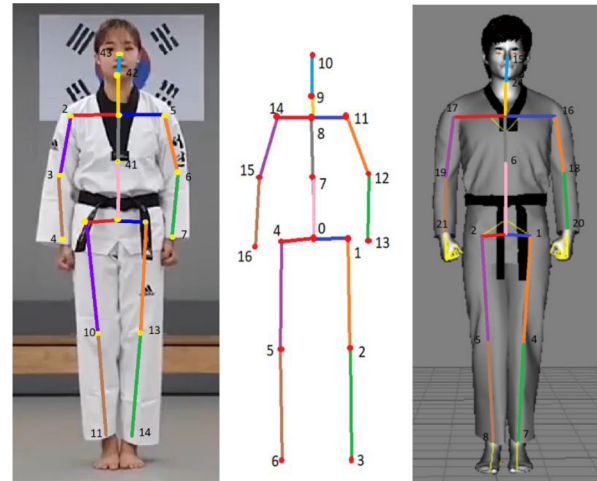
Taekwondo, such as high kicks and back-side views. Nevertheless, these poses did not dominate the sequences, and failures occurred only in specific joints.

The datasets and algorithms feature similar yet distinct joint hierarchies. Specifically, TCMR provides 49 outputs, including the 24 joints constituting the skin and an additional 25 joints inferred from skin position. In contrast, the reference animation encompasses 69 joints including fingers. Because no clear definition of joint locations was provided, we examined all joint locations, and selected the 17 joints (H36m) that are crucial for Taekwondo pose evaluation and considered almost identical in the reference and test data.



**Fig. 3.** Sample pose images from YouTube, including two professionals (a) and two practitioners (b)

To determine the positions of the left and right hips, we inferred the midpoints between their corresponding pairs. In addition, we computed the pelvis and neck as midpoints between the hips and shoulders, respectively (Fig. 4, Table 2). However, owing to the mismatch between the locations of the nose, head, and neck joints between the TCMR and reference data, we excluded these joints from the cost measurements. A joint local regression is required to compare more joints.



**Fig. 4.** Joint definitions: SMPL (left), H36m-17 (center), and FBX (right)

**Table 2.** Joint definitions: left SMPL, right FBX, center H36m-17

Joint name	TCMR	H36m	Reference
pelvis	midpoint of lhip and rhip	0	midpoint of 1 and 2
lhip, rhip	midpoint of 12 and 28, 9 and 27	1, 4	1, 2
lknee, rknee	13, 10	2, 5	4, 5
lankle, rankle	14, 11	3, 6	7, 8
spine, neck	41, midpoint of 5 and 2	7, 8	6, midpoint of 16 and 17
head, headtop	42, 43	9, 10	24, 15
lshoulder, rshoulder	5, 2	11, 14	16, 17
l elbow, relbow	6, 3	12, 15	18, 19
lwrist, rwrist	7, 4	13, 16	20, 21

**B. Results**

Fig. 5 depicts four prototypical instances of pose comparisons between the reference FBX and TCMR SMPL outputs, demonstrating that the proposed approach successfully detects specific differences in joint poses. Notably, the poses in Fig. 5(a), which are nearly identical, yield a mean pose error of 0.12. The poses in (b), (c), and (d) produced errors of 0.51, 0.60, and 0.27, respectively. The postures of the right and left legs in (b) differ significantly, and the postures of the left hand and both legs in (c) are also divergent. Despite the relatively small error value in (d), the pose error at joint 12 (left wrist) exceeds those of all other joints with a numerical value of 1.03 (Table 3).

Upon validating our pose comparison measures, we examined the performance of DTW synchronization. Fig. 6 presents a plot of matching frames between the synchronized sequences, and Fig. 7 shows human pose images across the sampled synchronized frames.

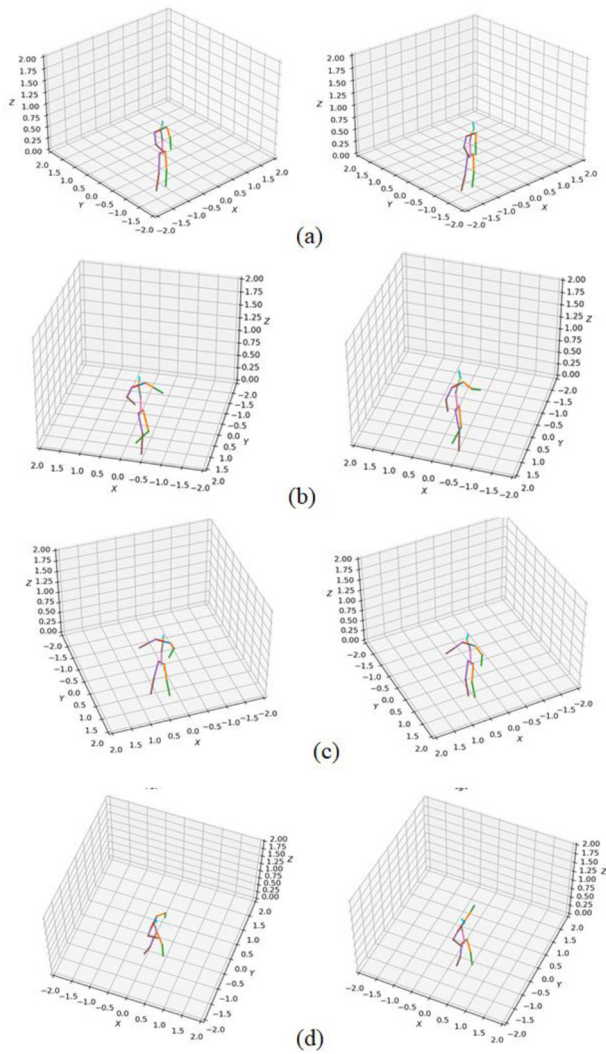


Fig. 5. Sample pose comparison examples

Table 3. Distance between joints of corresponding pose pairs

Joint ID	Fig. a	Fig. b	Fig. c	Fig. d
1	0.08	0.29	0.33	0.32
2	0.10	0.62	0.68	0.21
3	0.04	0.69	0.27	0.21
4	0.08	0.28	0.35	0.37
5	0.30	0.38	0.36	0.10
6	0.01	0.18	0.26	0.10
7	0.06	0.13	0.06	0.30
11	0.08	0.42	0.52	0.12
12	0.08	0.08	1.48	1.03
13	0.37	0.83	2.21	0.69
14	0.08	0.43	0.55	0.13
15	0.02	1.20	0.32	0.09
16	0.34	2.21	0.76	0.38

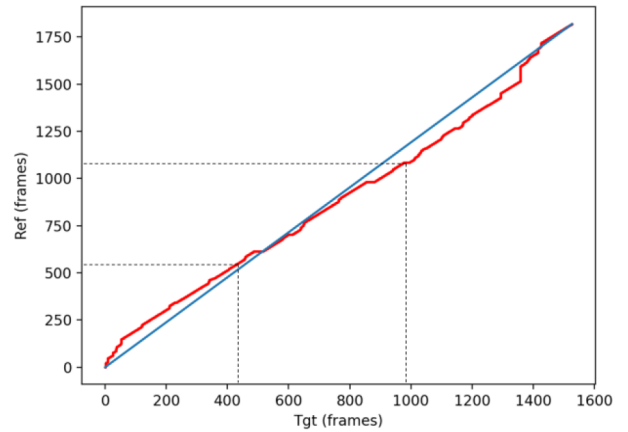
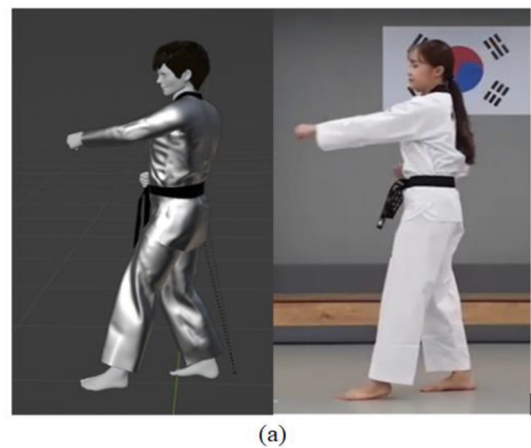


Fig. 6. Sequence synchronization (red line) between the reference (Ref) and target (Tgt)



(a)



(b)

Fig. 7. Two validating showcases of proposed pose sequence synchronization (at frame numbers (566, 450) on (a) and (1085, 993) on (b), where the left side is the reference and the right side is the target)

Following synchronization, the pose accuracies of the test videos were measured against reference pose sequences. Table 4 lists the overall pose accuracy results for Poomsae No.1 (Taegeuk 1, Jang), showing that the mean pose errors of professionals are smaller than those of practitioners, which correlates with the visual examination performed by Taekwondo experts. Fig. 8. depicts an example of score variation over time, demonstrating that our method effectively compares poses between professionals and practitioners. In Frame 430, the practitioner executed the Poomsae action correctly; however, in Frame 570, a mistake in the pose was made (see Fig. 9).

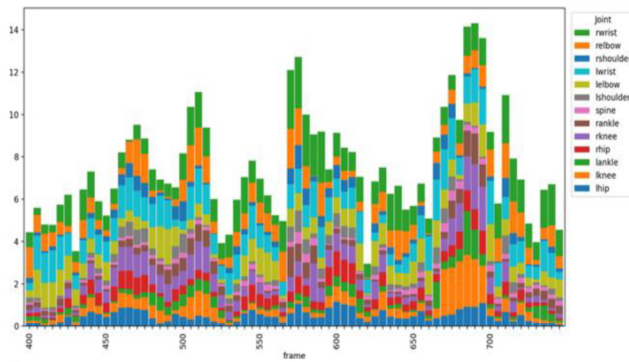


Fig. 8. Temporal stacked errors of poses

Table 4. Mean, minimum, and maximum cost values in performance between professionals (Pro) and practitioners (Prac), measured using our method

Reference	Target	Mean Cost	Min Cost	Max Cost
FBX	Pro_1	0.39	0.13	0.80
FBX	Pro_2	0.39	0.11	0.84
FBX	Prac_1	0.60	0.18	2.06
FBX	Prac_2	0.56	0.13	2.07

C. Discussion and Limitations

Although our method has shown promise in discriminating between professional and practical Poomsae performance, several aspects require further improvement for practical use. First, the accuracy of 3D HPE is crucial for future applications of our method. However, as shown in Fig. 10(a), the state-of-the-art monocular 3D HPE method often fails to approximate joint positions, especially for several unusual poses such as high kicks. In addition, as shown in Fig. 10(b), the estimation of the hands can be inaccurate in the back-side view owing to occlusion. A multi-camera method [1] must be deployed to overcome these inherent limitations. Despite these inaccuracies in pose estimation for many frames, our synchronization method matched the sequence sufficiently well to evaluate other frames.

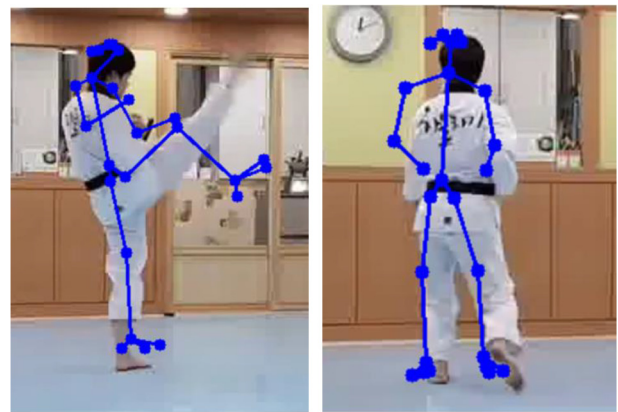


(a)



(b)

Fig. 9. Correct (a) and incorrect (b) action examples.



(a)

(b)

Fig. 10. Limitations of state-of-the-art 3D HPE method (a): unusual high kick pose, (b): occlusion from back-side view

Another issue is the potential mismatch between joint locations in the two models. Unfortunately, we could not obtain the definitions of all joints in the reference models; instead, we selected 17 joints that were visually similar and

shared the same joint names between the reference and TCMR models, and performed a simple regression. When joint definitions are available, a regression method can be applied from the TCMR joints to the reference joint locations. To examine differences between the joint definitions, we compared poses (TCMR joints) between professionals, with results listed in Table 5. The average costs were reduced by approximately 50% – from 0.4 to 0.2 – indicating that accuracy can be improved considerably when the joint definitions are matched. Specifically, joints 12 and 15, located in the upper part of the body, exhibited significant differences between the reference and TCMR models (Table 6).

**Table 5.** Comparison between reference and TCMR models

Reference	Target	Mean Cost	Min Cost	Max Cost
FBX	Pro_1	0.38	0.13	0.80
FBX	Pro_2	0.39	0.11	0.84
Pro_1	Pro_2	0.21	0.05	0.91

**Table 6.** Comparison between reference and TCMR models at initial (attention) pose

Joint ID	FBX vs. Pro_1	Pro_1 vs. Pro_2
1	0.06	0.13
2	0.07	0.01
3	0.10	0.07
4	0.06	0.13
5	0.28	0.22
6	0.03	0.05
7	0.04	0.07
11	0.03	0.22
12	0.59	0.08
13	0.22	0.05
14	0.04	0.22
15	0.56	0.27
16	0.26	0.10

#### IV. CONCLUSIONS

In this study, we designed an evaluation method to assess the performance of Taekwondo Poomsae practitioners. Our method addresses several challenges that arise when comparing martial art poses, including the lack of synchronization information and differences between human body shapes. We obtained shape-invariant pose measures and demonstrated that DTW matching can effectively synchronize pose sequences. The experimental results showcase the effectiveness of our method, highlighting its potential for supporting the development of Taekwondo Poomsae practitioners and enhancing their overall performance. Finally, it is important to note that an accurate 3D HPE is critical to our approach, and existing state-of-the-art monocular HPE techniques must

be improved. Future studies should explore more advanced algorithms for pose synchronization and develop more robust 3D HPE methods for challenging scenarios.

#### ACKNOWLEDGEMENTS

This study was supported by the Research Program funded by SeoulTech (Seoul National University of Science and Technology).

#### REFERENCES

- [ 1 ] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, “Deep 3D human pose estimation: A review,” in *Computer Vision and Image Understanding*, vol. 210, p. 103225, Sep. 2021. DOI: 10.1016/j.cviu.2021.103225.
- [ 2 ] K Kim, and Y. Kim, “Real-time dance evaluation by markerless human pose estimation,” *Multimedia Tools and Applications*, vol. 77, pp. 31199-31220, Jun. 2018. DOI: 10.1007/s11042-018-6068-4
- [ 3 ] R. Schramm, C. R. Jung, and E. R. Miranda, “Dynamic time warping for music conducting gestures evaluation,” *IEEE Transactions on Multimedia*, vol. 17, no. 2, pp. 243-255, Feb. 2015. DOI: 10.1109/TMM.2014.2377553.
- [ 4 ] Z. Cao, T. Simon, S-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, USA, pp. 7291-7299, 2017. DOI: 10.1109/cvpr.2017.143.
- [ 5 ] M. Kocabas, N. Athanasiou, and M. J. Black, “Vibe: Video inference for human body pose and shape estimation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, pp. 5252-5262, 2020. DOI: 10.1109/CVPR42600.2020.00530.
- [ 6 ] H. Choi, G. Moon, J. Y. Chang, and K. M. Lee, “Beyond static features for temporally consistent 3d human pose and shape from a video,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Nashville, USA, pp. 1964-1973, 2021. DOI: 10.1109/CVPR46437.2021.00200.
- [ 7 ] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A skinned multi-person linear model”, *ACM transactions on graphics (TOG)*, vol. 34, no. 6, pp. 1-16, Nov. 2015. DOI:10.1145/2816795.2818013.
- [ 8 ] K. Isakov, E. Burkov, V. Lempitsky, and Y. Malkov, “Learnable triangulation of human pose,” in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, Seoul, Korea, pp. 7717-7726, 2019. DOI: 10.1109/ICCV.2019.00781.
- [ 9 ] S. Chun, S. Park, and J. Y. Chang, “Learnable human mesh triangulation for 3D human pose and shape estimation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, pp. 2849-2858, 2023. DOI: 10.1109/WACV56688.2023.00287.
- [ 10 ] H. Dai, B. Cai, J. Song, and D. Zhang, “Skeletal animation based on BVH motion data,” in *2010 2nd International Conference on Information Engineering and Computer Science*, Wuhan, China, pp. 1-4, 2010. DOI: 10.1109/ICIECS.2010.5678292.
- [ 11 ] T. Hachaj, M. R. Ogiela, and K. Koptyra, “Application of assistive computer vision methods to Oyama karate techniques recognition,” *Symmetry*, vol. 7, no. 4, pp. 1670-1698, Sep. 2015. DOI: 10.3390/sym7041670.
- [ 12 ] KCISA, Taekwondo Poomsae 3D data, 2020. [Online], Available: <https://www.kcisa>.





**Thi Thuy Hoang**

Thi Thuy Hoang was born in Vietnam, in 1994. She received the B.S degree in Control and Automation Engineering from Le Quy Don Technical University, Hanoi, Vietnam, in 2018. Now, she is working toward the M.S degree in Electrical and Information engineering at SeoulTech. Her current research interests include computer vision, machine learning, and human pose estimation.



**Heejune Ahn**

Heejune Ahn was born in Seoul, Republic of Korea, in 1970. He holds a Bachelor of Science (BS), Master of Science (MS), and Doctor of Philosophy (Ph.D.) degrees, all earned from KAIST, Republic of Korea, in 1993, 1995, and 2000, respectively. He completed his postdoctoral research at the University of Erlangen-Nuremberg, Germany in 1999. He also has professional experience as a senior engineer at LG Electronics and as a chief engineer at Tmax Soft Inc, South Korea. In 2004, he joined SeoulTech, Republic of Korea, and has been a full professor since 2013. He has also worked as a visiting professor at UVa, USA in 2011, and Cardiff University, UK in 2019. His research interests include computer vision, machine learning, and computer networks.