PoomsAI: An Accuracy Auto-grader Using Bottom-up Human Pose Estimation

Kevin Jiang
Massachusetts Institute of Technology
305 Memorial Dr, Cambridge, MA 02139
kev2018@mit.edu

Joshua Lee
Massachusetts Institute of Technology
305 Memorial Dr, Cambridge, MA 02139
jlee2022@mit.edu

Abstract

We present PoomsAI - a novel automated scoring system for Taekwondo Sport Poomsae that achieves similar or better performance accuracy scores by human judges. Currently, both accuracy and presentation sections in a performance are graded by humans, who are easily error-prone due to the number of criteria to watch. Our goal is to make judging easier and more correct by making the accuracy section grading automatic. To our knowledge, there currently does not exist any competing methods that accomplish this task. Our approach consists of taking video footage of a competitor's performance, determining the competitor's pose through keypoint identification, and then performing vector analysis on the 2D pose estimation to provide accuracy scores. Experiments show that our model performs on par with human judging on mediocre-to-good performances.

1. Introduction

Taekwondo, a Korean martial art, consists of two main divisions in the national and international competitive scenes: Kyorugi and Poomsae, which are commonly known as sparring and forms, respectively [6]. While Taekwondo sparring is currently present in the Olympic scene, Taekwondo poomsae is not, but it is being advocated for. In this paper, we will be focusing on poomsae/forms, which is a defined routine consisting of a set of hand and feet movements that an athlete has to perform, instead of sparring. Poomsae contains multiple divisions divided primarily by age group. For each division, there is a set of fixed routines that every athlete has to know and prepare.

At a poomsae competition, competitors perform in rings and are judged by a set of five or seven judges [6]. The general layout of a competition ring is depicted in Figure 1. When it is time for a competitor to perform a form or routine, they step onto the ring and begin when the coordinator gives the signal. Upon completing their routine, the competitor then exits the ring after the coordinator's signal and the judges begin inputting their scores into the system. Table 1 shows the distribution of points, where the maximum score is 10, and the criteria for both the accuracy and presentation sections. At the start of a performance, a perfect accuracy score of 4.0 is assumed. For each incorrect movement, either a 0.1 or 0.3 deduction is made from the score depending on the severity. Each judge submits their own score for the performance's accuracy and presentation (e.g. 3.3 for accuracy and 4.2 for presentation). Before the score for a competitor is calculated, the highest and lowest score for both accuracy and presentation is dropped. The final score for a competitor is the sum of the accuracy average and presentation average after the dropping [7]. This process repeats until the division finishes, and the winner is the competitor with the highest score.

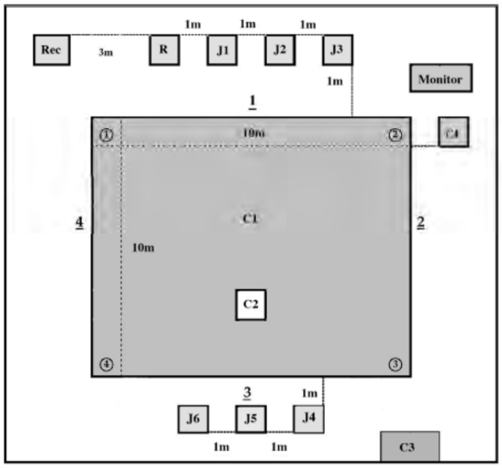


Figure 1. The illustration of a poomsae competition ring. Human judges are depicted as R, J1, J2, etc., where R represents the referee. The grey square (C1) consist of the competition area, and the athlete begins and ends their performance on C2. The athlete's coach stays at the bottom right of the ring (C3). The competition coordinator stands on the top right (C4). If a recorder is used, it is placed on the top left (Rec). Competition scores and results are displayed on a computer screen (Monitor) [6]

Scoring Criteria	Details of Scoring Criteria	Points
Accuracy (4.0)	Accuracy of details of each Poomsae	4.0
	Accuracy of basic movements and balance	
Presentation (6.0)	Speed and power	2.0
	Strength/speed/rhythm	2.0
	Expression of energy	2.0

Table 1. Scoring Criteria for Poomsae

While presentation scoring is highly subjective due to the difficult nature of quantifying stylistic preferences like ”expression of energy” and ”rhythm”, accuracy is more objective since moves are strictly defined and categorized (e.g. in a walking stance, the length is three feet long from the origin, the back foot is turned 30 degrees, the legs are both straight, etc.) [6, 7]. A typical routine contains 20 to 30 moves and lasts anywhere from 30 seconds to just over one minute. This speed makes judging both accuracy and presentation difficult and error-prone. By creating an automated system to grade performance accuracy, we alleviate the amount of work for judges and allow them to concentrate on the presentation portion.

2. Related Work

There have been previous attempts to incorporate advances in computer vision to improving sports assessment, including ball tracking in baseball and Hawk-Eye in tennis[5]. Both systems depend on real-time visualization data from cameras located in various parts of the field or court to infer information about the ball, including velocity, spin, and overall trajectory. These advancements have been utilized to assist traditional referees and serves as a backup to increase impartiality in sports.

More recently, Fujitsu, a Japanese company, has developed an automated judging system for gymnastics that uses multiple camera angles to determine skeleton points in the gymnast, then evaluating the performance based on the physical accuracy, including angles held and relative locations of certain body parts [2]. This support system is intended to be used as an assistant to the judges for the Tokyo 2020 Olympics.

There have also been previous efforts to evaluate kicks in taekwondo using sensors that were attached to the protective gear of the athlete. This accurately measures the strength and speed of the kick, which is an important aspect of the poomsae, but lacks evaluation for other parts considered in grading. In our project, we attempt to isolate ourselves from the need of extra equipment and rely solely on visual data to assess the performance of the athletes.

The method used in our grader, which uses OpenPose to generate keypoints in various parts of the body for each frame, has been employed in other projects such as ”Everybody Dance Now” by a group of students from UC Berkeley [1]. This software uses the poses extracted from an original video to transfer to a target subject, thereby creating an identical rendition of the dance performed by the target subject.

3. Approach

Our approach to create an automated accuracy scoring system can be broken down into three main components: attaining video footage of the routine, finding a mapping from the raw footage to pose estimates, and using the pose estimates to perform vector analysis to determine accuracy scores. In this paper, we attempted to grade the most basic poomsae called Taegeuk Il Jang, a form that all white belts start with. We simplified the number of parameters to be dealt with by using only the front-facing camera angle instead of capturing multiple angles. Moreover, we assume all mistakes will cause a 0.1 deduction.

3.1. Footage

To attain video footage, we used a regular Apple iPhone XS camera to film multiple performances of the poomsae by different genders and in different tiers of quality: good, mediocre, and bad. A ”good” performance is defined as nearly all elements were performed accurately, ”mediocre” is defined as a good performance with several incorrect moves, and ”bad” is defined as a majority of movements are mistakes or performing a completely different form.

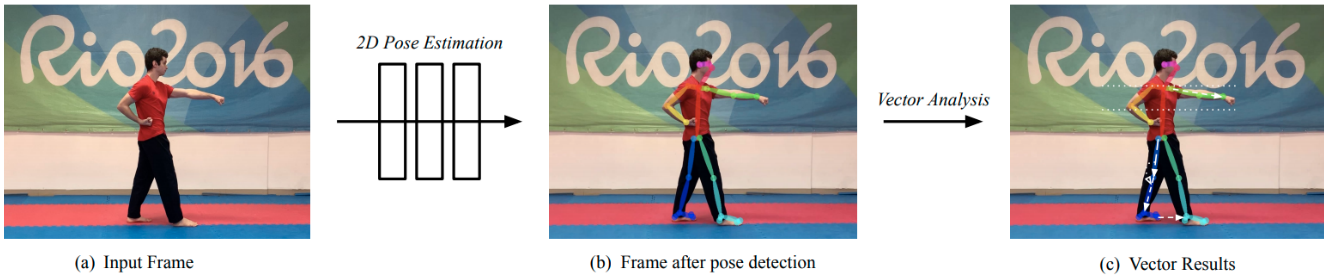
3.2. Finding Pose Estimates

Given video footage, we extrapolated relative distances and angles between various body parts of the athlete to judge accuracy. We observed that using keypoints preserves motion signatures over time while removing as much extra information as possible. One way of extracting keypoints in an image is using a human pose estimation algorithm, which identifies the localization of human joints. In this case, we used a human pose detector called OpenPose to provide a mapping from raw video footage to generate pose estimates for each frame. OpenPose uses a convolutional neural network (CNN) for body part detection and part affinity fields (PAFs) for part association [3].

3.3. Vector Analysis

After obtaining 2D pose estimates for each frame of the video footage, we generated accuracy scores. We split the scoring into two parts: transitions and ”stopping points”. Stopping points are the times when a particular move finishes (i.e. each move has a start and ending point). By detecting the stopping points throughout the performance, the algorithm can use the frames in between to determine transition accuracy.

Because Taeguk Il Jang follows a very specific sequence of movements, we identified 20 poses that the athlete would



(a) Input Frame

(b) Frame after pose detection

(c) Vector Results

Figure 2. Overall analysis pipeline. (a) After identifying a ”stopping point”, the video frame becomes the input to a CNN that performs (b) human pose detection . (c) Using the 2D coordinates of the keypoints in the output, we analyze the relation of points/vectors to assess if the move meets the accuracy standards.

in theory hold momentarily. In a good performance, the athlete would give sharp indications of a stop, and the overall accuracy could be measured by evaluating each of these 20 stops. In the video footage, stopping points were initially filtered by measuring the velocity of each body part, determined by the change in position between frames. Body parts that are considered more indicative of a stop, including arms, legs, and key joints were weighted more heavily in the consideration of velocity. Frames where the magnitude of the velocity did not reach a magnitude of 0.625 pixels per frame were considered to be stopping points, and consecutive frames where this held were considered to be the frames where the same stance had been held over. The middle frame was then sent forward to be used as the model to evaluate the quality of the pose for that stopping point. To accommodate for poor performances where it is impossible to detect a stopping point based on just velocity, a measure of acceleration for arms and legs was implemented by finding the difference in velocity between three consecutive frames. A high acceleration indicates a change in motion, which is a good indicator that the athlete proceeded to the next move. In this case, the middle frame of the three was considered to be the stopping point and was used for the next part.

For each stopping frame, we have the keypoint coordinates outputted by the 2D pose estimation. Using these coordinates, we generated the corresponding vectors to grade the technique(s) and stance(s). Figure 2 depicts the overall approach to the algorithm. In the picture, the athlete performed a left middle punch with a right walking stance. Therefore, in this instance, criteria such as having a straight arm for the punch can be determined easily by performing the calculation

$$\theta = \arccos\left(\frac{\vec{v_1}\cdot\vec{v_2}}{\|\vec{v_1}\|\|\vec{v_2}\|}\right)$$

where $\vec{v_1}$ and $\vec{v_2}$ are the vectors from the elbow to the wrist and the elbow to the shoulder, to get an estimate on the arm angle of the athlete. Moreover, we graded the walking stance on criterion such as stance length, which can be estimated by finding the magnitude of the vector from the left foot's toe to the right foot's heel. If any of the defined criterion for each move is not met, the move is marked as incorrect, and the algorithm deducts 0.1 points. The algorithm analyzes all of stopping frames and subtracts all of the deductions from 4.0 to obtain a predicted accuracy score. More specifically, we graded each hand technique, kick, and stance based on the guidelines outlined in the USA Taekwondo Poomsae Competition Rules document [7].

As an example, here are the detailed criterion we used to grade hand techniques. We created a threshold algorithm to judge middle punches, low blocks, inside blocks, and high blocks. Middle punches had four main criterion: the punching arm needs to be straight, the punch's height should be near the solar plexus, the opposite hand needs to be placed correctly on the sides, and the punch needs to be at the center-line of the body. Low blocks also had four criterion: the blocking hand should be straight and two fist widths above the quadriceps, and the opposite hand should be placed correctly. For inside blocks, the blocking arm should be bent 90 to 120 degrees with the wrist between the center and shoulder level, and the opposite hand should be placed correctly. Finally, for high blocks, the blocking arm should be one fist width from the forehead at a 45 degree angle with the opposite hand placed correctly. For the specific algorithm and calculations we performed to grade the criteria and the complete list of criterion we used for stances and kicks, refer to our GitHub [4].

4. Experiemental Results

After building the automated scoring system, we tested the performance of the auto-grader by comparing the accuracy results to real human judging scores. We took three individuals and asked them to perform with the three tiers of quality - good, mediocre, and bad. For each of the nine video footage, we asked a set of three referees for their respective accuracy scores and took the average to generate the final human judge score per performance. To simulate a

	Human Judges	PoomsAI	Score Difference
Individual 1 (Good)	3.6	3.3	-0.3
Individual 2 (Good)	3.5	3.4	-0.1
Individual 3 (Good)	3.4	3.6	+0.2
Individual 1 (Mediocre)	2.8	2.7	-0.1
Individual 2 (Mediocre)	2.9	2.7	-0.2
Individual 3 (Mediocre)	2.7	2.7	0.0
Individual 1 (Bad)	1.6	0.2	-1.4
Individual 2 (Bad)	2.0	1.1	-0.9
Individual 3 (Bad)	1.9	1.5	-0.4

Table 2. PoomsAI Performance vs. Human Judges

real competition environment, we asked the judges to give a score after viewing the performance once. The following describes our results for the stopping points detection algorithm and PoomsAI accuracy results.

The algorithm for determining stopping points was mostly successful in returning the relevant frames. Across the three individuals for which the model was tested upon, anywhere from 16 to 18 out of the 20 stopping points were successfully determined for the mediocre and good performances, with no false positives. Remaining stopping points were simply not included in the overall grade, although a missing stopping point was negatively reflected. For poor performances, the algorithm was only able to determine 11 to 15 of the 20 stopping points correctly, although no false positives were returned similarly to the case with better performances. This reflected negatively on the score given to the poor performances by the grader, as shown below in Table 2.

Using the frames identified by the stopping points algorithm, the auto-grader determined if each the movements of a particular performance met the corresponding criteria. Table 2 displays the results of the PoomsAI system and human judging for the nine different performances.

Nearly all of the PoomsAI scores were lower than the human judges' because our scoring system noticed details in movements that a human might miss given the speed of the movements. Furthermore, for the good and mediocre performances, the auto-grader gave relatively similar scores with the largest margin being 0.3. The bad performances, however, had large differences, with our system outputting much lower scores than the judges. This result is mostly due to a lack of identification of stopping points. The bad performances tended to connect many movements together without ever pausing. As a result, the stopping algorithm failed to detect a valid movement from the athlete, and the auto-grader assumed that the movement was completely incorrect.

5. Conclusion

PoomsAI is able to correctly judge performances that would be considered mediocre or better, defined roughly as performances that would be awarded 2.5 points or better by a human judge. The discrepancy between bad and good performances comes from the difficulty in correctly identifying stopping points for videos where the athlete does not produce concrete queues to indicate a stop.

Limitations to our model include only being able to judge Taeguk Il Jang, the most basic form. In addition, the model was not able to grade enough videos for us to confidently propose this system as a viable replacement for human judges, due to insufficient data.

Further work on PoomsAI may involve

1. Improving the stopping point finding algorithm
2. Using multiple camera angles to extrapolate 3D coordinates of the athlete
3. Using a neural network to translate stop frames into scores, trained from data from collegiate and national tournaments
4. Expanding grader to accommodate other forms
5. Incorporating quality of transitions (frames between stopping points) into overall grade.

We believe that our current approach is promising, and using multiple camera angles to develop a 3D model at each frame would allow us to perform much more accurate calculations. We will also explore the option of developing a convolutional neural network to automatically identify specific moves and whether or not they are correct. Creating such a general network would make judging new forms easier, but it would require many images to train. One possibility of getting the amount of images needed to train the model would be to take images and footage at collegiate or national competitions.

6. Contributions

Kevin obtained video footage of different performances for testing and training. Josh's work includes developing the algorithm for the stopping points, and both Kevin and Josh worked on incorporating OpenPose to yield videos that could be worked with. Kevin developed the algorithm to process stopping points to yield a grade for the athlete. In addition, Josh wrote the abstract, related work, and Kevin and Josh both wrote the approach, the experimental results, conclusion, and references. Kevin wrote the introduction. Both Josh and Kevin also made the presentation that corresponds to the parts each of them worked on for the paper.

References

[1] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros. Everybody dance now, 2018.

[2] H. Fujiwara. Ict-based judging support system for artistic gymnastics and intended new world created through 3d sensing technology. *Fujitsu Science and Technology Journal*, 54(4):66–72, 2018.

[3] G. e. a. Hidalgo. Openpose, 2019. GitHub Repository, https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[4] J. L. Kevin Jiang. Poomsai, 2019. GitHub Repository, https://github.com/p1ck-4-u53rn4m3/poomsae-grader.

[5] C. S. N. Owens, C Harris. Hawk-eye tennis system. *Visual Information Engineering*, 1(495):182–185, 2003.

[6] U. Taekwondo. Usa taekwondo poomsae athletes' reference guide, 2010.

[7] U. Taekwondo. Usa taekwondo poomsae competition rules, 2016.