Development of Parallel Stance Check System for Ski Videos Using Pose Estimation Model

Hiromu Ishiwata s1300109 Supervised by Prof. Ken Nakazawa

Abstract

Parallel Stance in skiing is an important pose in which both feet are kept parallel during a parallel turn, with both skis parallel to each other. Until now, parallel stance quality has been determined by subjective human evaluation, resulting in issues with subjectivity and reproducibility. Therefore, in this study, I developed parallel stance check system for checking the quality of parallel stance from ski videos. The proposed method uses ski videos shot from below or behind, and targets are detected using YOLO by reverse playback and ROI design according to the shooting direction to stabilize target detection. After comparing and evaluating 2D and 3D pose estimation models, I selected 2D pose estimation model and performed to check parallel stance based on angle between both shins. The angle threshold was determined using parallel & not parallel angle distribution histogram created from subjectively labeled frames. Furthermore, I applied judgement label smoothing using a median filter to smooth out incorrect pose estimation model estimates. Results confirmed that the proposed method improved target detection stability and enabled appropriate judgments based on the skier's form in many frames. I also demonstrated that smoothing the judgment labels suppresses incorrect judgments caused by temporary incorrect pose estimation, improving the stability of parallel stance check. In conclusion, this study's method is effective in improving the objectivity and reproducibility of checking parallel stance using ski videos, and is considered useful as a method to support coaching and independent practice. However, issues remain regarding the accuracy of judgments, such as decreased accuracy depending on the shooting method and repeated incorrect pose estimation. In the future, I hope to develop a more general-purpose parallel stance check system by introducing 3D pose estimation model and improving the judgment criteria.

1 Introduction

In skiing, parallel turn is a fundamental and important turning technique performed while maintaining both skis parallel. Parallel Stance (PS) refers to the pose in which both feet are kept parallel during the parallel turn [1]. PS quality has traditionally been determined primarily through subjective human assessment. However, visual assessment tends to rely on experience and subjectively, resulting in inconsistent assessments even for the same run. Furthermore, it is difficult for skiers to objectively assess their form from their ski videos, making it difficult to quantitatively grasp the extent to which PS is maintained at each stage. Thus, checking PS face challenges in terms of reproducibility and objectivity. Meanwhile, recent advances in deep learning have significantly improved human pose estimation from images and videos. Furthermore, attempts have been made to analyze skiing movements using pose estimation model in related skiing research [2][3]. However, no research has been conducted aimed at automated to check PS. However, by using pose estimation technology, it may be possible to automatically perform to check PS and objectively evaluate form. Therefore, in this study, I developed a system that automatically perform to check PS from ski videos. The purpose of this study is to create a system that automate to check PS in an objective and reproducible manner and can be used practically in technical instruction by instructors and in the self-practice of skiers themselves. By objectively evaluating to check PS, which previously relied on experience and subjectivity, I aim to efficiently support skiers in understanding their technique and improving it.

2 Methods

2.1 System Overview

First, I considered the general operation of the system. The process shown in Fig. 1. is repeated for all frames of the input ski videos.

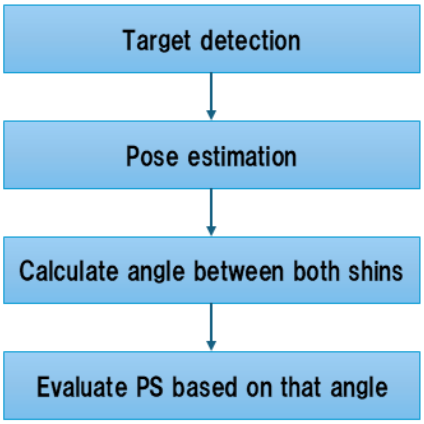

Fig. 1 System Overview

2.2 Problem Definition

When developing PS check system,
the following challenges are likely to arise:

- Which object detection model to use
- How to track target
- Which pose estimation model to use
- How to perform to check PS based on angle between both shins (angle).
- How to deal with poses incorrectly estimated by the pose estimation model

I aim to resolve these challenges and build a system that can easily perform to check PS on ski videos.

2.3 Considerations and Solutions

2.3.1 Ski Videos to Upload and Shooting Method

I determined to use ski videos that contain skiers who are target of checking PS, and that are shot from below or behind, which makes pose estimation relatively easy, regardless of the estimation dimension of the pose estimation model.

2.3.2 Playback Direction Depending on Shooting Method

For ski videos shot from below, I play them in reverse playback and ask the user to manually enter where target are in the last frame. For ski videos shot from behind, I play them playback and ask the user to manually enter where target are in the first frame.

2.3.3 How to Track Target

For the object detection model, I considered YOLO's yolo12n.pt [4][5], which is highly accurate and fast. ROI is designed by multiplying the diagonal of the YOLO bounding box (bbox) detected in the previous frame by 1.5 to detect and track the target.

2.3.4 Consideration of Pose Estimation Model

I compared and examined yolo11x-pose.pt of YOLO [6], which performs 2D pose estimation, and Blaze Pose's pose_landmarker_full.task [7][8], which can estimate 3D pose.

2.3.5 How to Perform to Check PS Based on Angle

The threshold is determined in the following way.

① Obtain hundreds of skier frames.
② Subjectively judge whether it is parallel or not parallel and label it.
③ Estimate pose of all frames using the pose estimation model and calculate angle.
④ Create histogram that aggregates how many parallel and not parallel each angle contains.
⑤ Determine threshold from histogram.

2.3.6 Smoothing to Judgement Labels

Since pose estimation model may sometimes make an incorrect pose estimation, median filter is applied judgement labels to smooth the time series. Specifically, if the judgments of the previous and next frames match for three consecutive frames, and the judgment of only the center frame is different, the judgment of the center frame is considered to be the same as the previous and next frames.

3 Results

In this chapter, I show the effectiveness of the proposed target detection method, which pose estimation model to use, the determination of the angle threshold for checking PS, and the effects of judgement labels smoothing.

3.1 Effectiveness of Target Detection Using Reverse Playback and ROI

I conducted target detection test to verify whether the proposed method could efficiently detect target. Fig. 2. shows comparison of target detection test using yolo12n.pt for a ski video shot from below using playback & no ROI (Normal) and the proposed method. Confidence score 0 indicates a case where target detection failed. In the normal method, the confidence score fluctuates between 0.6 and 0.8, and target detection fails around 16 to 18 seconds of the ski video. On the other hand, proposed method maintains confidence score above 0.8, also indicating that target detection is successful throughout. Also, from Fig. 3, I can confirm that using the proposed method rather than Normal results in higher scores for average confidence score and target detection success rate. Based on these results, I determined to use the proposed method for target detection.

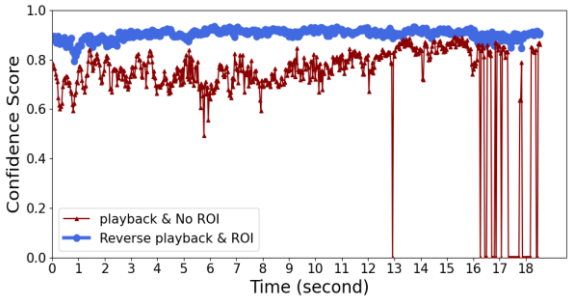

Fig. 2 Comparison of target detection test using playback & No ROI and Reverse playback & ROI

	Average confidence score	Target detection success rate
Proposed method	0.904	99.82 %
Normal	0.77	93.53 %

Fig. 3 Average confidence score and target detection success rate of proposed method and Normal

3.2 Determine Which Pose Estimation Model to Use

To determine which pose estimation model to use, Fig. 4. compared the accuracy of pose estimation between pose_landmarker_full.task and yolo11x-pose.pt. Since Blaze Pose does not display confidence score like YOLO, I averaged all joint point scores and displayed them in the upper left. While pose_landmarker_full.task was unable to estimate the correct pose, yolo11x-pose.pt was able to estimate the correct pose. Therefore, I determined yolo11x-pose.pt as pose estimation model in this study.

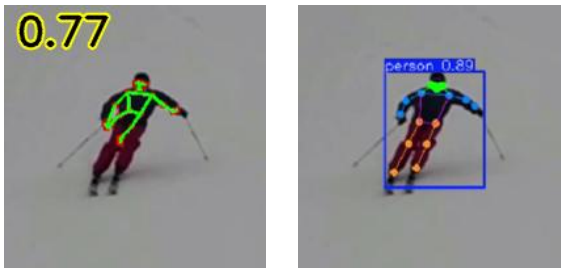

Fig. 4 Comparison of accuracy of pose estimation between pose_landmarker_full.task (left) and yolo11x-pose.pt (right)

3.3 Angle Threshold for Checking PS

To perform to check PS based on angle, I create a histogram of Fig. 5 as described in 2.3.5 and determine the threshold from it. The parallel ratio for all frame in one section is also plotted as line graph. Green line is 50% line of parallel ratio. These results, angle threshold is:

angle < 14 ➡ parallel
14 <= angle < 16 ➡ controversial form
16 <= angle ➡ not parallel

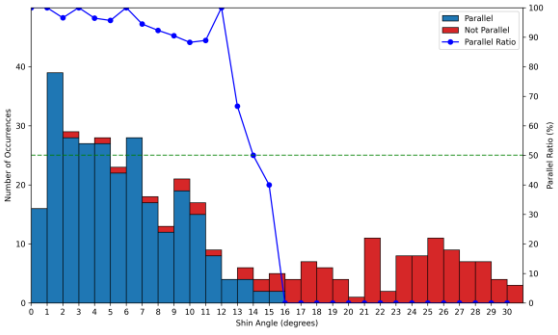

Fig. 5 Histogram of parallel & not parallel distribution for each angle and line graph of parallel ratio

3.4 Smoothing to Judgement Labels

To confirm the effect of smoothing to judgement labels, I compared original and median filtered of judgement labels. Fig. 6. shows comparison of original and median filtered of judgement labels from a ski video of not parallel skier. Even though the skier was not actually parallel, there were moments during

the turn when the skier was shot beside them, causing yolo11x-pose.pt to mistakenly estimate the pose as parallel. Therefore, before median filtering, judgement labels is parallel around 1.1, 1.2, 1.5, and 1.9 seconds. If I leave this, the system would mistakenly judge the skier was parallel, even though the skier was not actually parallel. After median filtering, judgement labels was smoothed, and checking PS was correct for the actual form in frames other than around 1.2 seconds. However, smoothing was not possible in the frame around 1.2 seconds where a series of incorrect pose estimation occurred. However, since it is effective for detecting a single incorrect pose estimation, I will smooth the judgment labels.

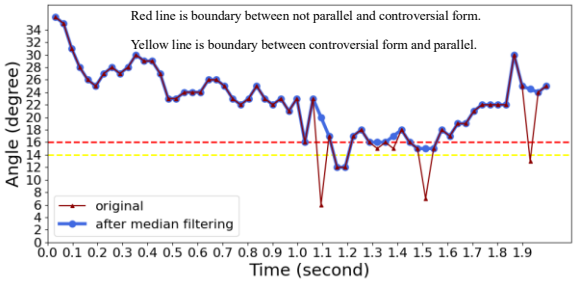

Fig. 6 Comparison of original and median filtered of judgement labels

3.5 Flowchart of PS Check System

Based on the all results, flowchart and specifications of PS check system and are as follows:
The process in Fig.7. is repeated for all frames from input ski videos.

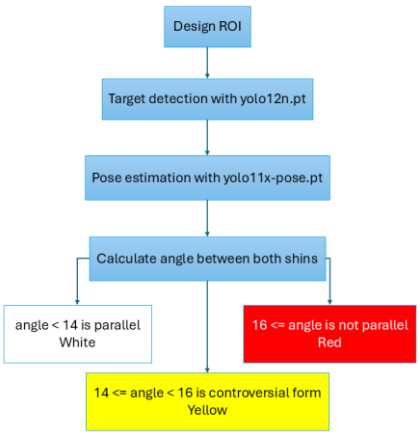

Fig. 7 Flowchart of PS check system

Once all frames have been processed, judgement labels is median filtered and video is presented to the user as the result of checking PS.

4 Discussion

4.1 Discussion of Results

4.1.1 Method for Target Detection Using Reverse Playback and ROI

It is believed that yolo12n.pt is able to recognize targets more easily, resulting in an increase in confidence score and target detection success rate.

4.1.2 Determination Pose Estimation Model

yolo11x-pose.pt is excellent not only in pose estimation but also in object detection, which is why it was able to recognize target even if they were located a little far away. On the other hand, Blaze Pose is lightweight for mobile devices and is designed to assume that input is from people who are very close to the camera. Therefore, if the target is located even slightly far away, it cannot be accurately recognized, and pose estimation inevitably fails.

4.1.3 Histogram of Parallel & Not Parallel Distribution for Each Angle

In this study, the form was subjectively judged as either parallel or not parallel, which is thought to be why the results are as shown in Fig. 3. Also, there are not parallel angles between 2 and 13 degree, but these are frames where yolo11x-pose.pt mistakenly estimated a parallel pose, so angle is also calculated within the parallel range.

4.1.4 Smoothing to Judgement Labels

If correct pose estimation had been performed before and after, it is thought that the incorrect pose estimation that occurs in the middle frame could have been smoothed out using median filter.

4.2 Effectiveness and Limitations of the Proposed Method

4.2.1 Method for Target Detection Using Reverse Playback and ROI

By playing ski video shot from below in reverse, yolo12n.pt can more easily recognize the target, and it can track the target in subsequent frames by designing a ROI based on a diagonal line. However, if the target is located too far away, yolo12n.pt can't

recognize it using the ROI, and tracking is interrupted.

4.2.2 Accuracy of 2D Pose Estimation Model yolo11x-pose.pt

In this study, I adopted yolo11x-pose.pt because 3D pose estimation model pose_landmarker_full.task was unable to accurately estimate pose of skier. However, the disadvantage of 2D pose estimation remains an issue. For example, even if the pose is not parallel when viewed from the front, it may be incorrectly estimated as parallel when viewed from the side. Also, if the target is located too far away, the image quality will be poor and pose will be more likely to be misestimated, which inevitably makes it difficult to perform correctly to check PS.

4.2.3 Determine Threshold of Angle

In determine threshold, labeling is based on subjectivity. Therefore, there will likely be some mismatch between boundary between parallel and not parallel to other people. This means that there is inevitably possibility that there will be mismatch between the judgment of everyone and the judgment based on angle threshold of system.

4.2.4 Smoothing to Judgement Labels

As shown in Fig. 5. , if an incorrect pose estimation occurs once, it can be smoothed out with median filter. However, if it occurs repeatedly, the median filter cannot filter it out, and it may make an incorrect judgment such as parallel when in fact it is not parallel.

4.3 Practical Conditions

From the above, it can be said that this is a practical system that can perform accurately to check PS if the target is skiing not far from the cameraman and the ski videos is recorded using the specified shooting method.

4.4 How Can It be Used for Teaching and Self-Practice

This study did not involve having actual users try it out to verify its effectiveness. However, if this system is realized, it will automatically check PS and visualize at which point during performance (first half/second half, right/left turns, etc.) PS is not being maintained. This will allow instructors to objectively share points to correct during instruction and enable skiers to quickly identify their issues and bad habits during self-practice.

5 Conclusion

5.1 What I Did in This Study

In this study, I developed system that automatically check PS from ski videos so that instructors and skiers practicing can easily check PS.

5.2 What was Achieved

The proposed method used reverse playback and ROI to improve the target detection success rate and confidence score so that yolo12n.pt could efficiently recognize targets from ski videos. In addition, by focusing on angle from the pose estimated by the pose estimation model yolo11x-pose.pt, confirmed that it was possible to make appropriate judgments based on the skier's form in many frames.

5.3 Future Challenges

However, because this method is based on 2D pose estimation, there remains the issue of reduced accuracy in judgments when the ski video is shot from the side or when the target is located too far away. Furthermore, if incorrect pose estimations occur repeatedly, smoothing with median filter is not possible, which may result in incorrect judgments. Future work includes introducing a new 3D pose estimation model, relaxing the shooting method, and further improving the judgment criteria, which are expected to lead to the development of a more versatile PS check system.

Reference

[1] 佐藤政之, スキー指導の進め方-シュテムターンからパラレル・ターンまで, 駒澤大學 北海道教養部論集 7, pp. 27-37, 1992.

[2] Martinelli G., Diprima F., Bisagno N., Conci N., Ski Pose Estimation, IEEE, 2024 IEEE International Workshop on Sport, Technology and Research (STAR), pp. 120-125, 2024.

[3] Bao W., Niu T., Wang N., Yang X., Pose estimation and motion analysis of ski jumpers based on ECA-HRNet, Springer Nature, Scientific Reports 13(1), Article 6132, 2023.

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., You Only Look Once: Unified, Real-Time Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16), pp. 779–788, 2016.

[5] Ultralytics, YOLO12n [Software], GitHub, 2024. Available : https://github.com/ultralytics/ultralytics

[6] Ultralytics, YOLO11x-Pose [Software], GitHub, 2024. Available : https://github.com/ultralytics/ultralytics

[7] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M., BlazePose: On-device Real-time Body Pose Tracking, arXiv preprint arXiv:2006.10204, 2020.

[8] Google AI Edge, Pose Landmarker (Full Task), Google, 2023. Available:https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker