APE-V: Athlete Performance Evaluation using Video (Supplemental Material)

Chaitanya Roygaga[1], Dhruva Patil[2], Michael Boyle[2], William Pickard [2], Raoul Reiser[2], Aparna Bharati[1], Nathaniel Blanchard[2]
[1] Lehigh University, PA, USA [2] Colorado State University, CO, USA

1. Dataset

1.1. Countermovement Jump

The countermovement jump [7] is a simple and reliable measure of lower-body power. The jump helps coaches in determining performance changes [11] and fatigue levels. Performances in the countermovement jump are linked with maximal speed, maximal strength, and explosive strength. The countermovement jump consists of three phases [7], which are performed without pausing. It begins with the participant standing straight on the force plate, with eyes in the front. The participant then undergoes an unweighting phase, after which there is a braking phase. At this point, the participant's knee angle is at about 90 degrees, which is subject to each individual. A successive propulsive phase puts the participant in the flight phase, while fully extending their legs and using the momentum to jump higher. The athlete then lands as close to the jumping off point as possible, thus completing one complete motion of the jump [see Figure 1].

1.2. Drop Jump

The drop jump [2] is designed to examine athlete reactivity. It is considered a fast stretch-shortening movement [9]. One of the main measures of the test is how quickly the athlete can move from absorption to propulsion. It also provides a qualitative indication of an athlete's lower limb alignment [8] in the frontal plane, with a straightforward drop-jump and subsequent and immediate vertical jump. The drop jump has phases similar to the countermovement jump. It begins from an elevated platform. The first phase is the drop phase, in which the participant jumps onto the force plate. As the athlete drops to the ground, they should land with their knees slightly bent. This is the braking phase, also called as the deceleration phase. The propulsive, flight and landing phases are similar to the countermovement jump [see Figure 1].

1.3. Errors in jump motion

The following rules have been used to annotate the errors [Refer Table 1 for complete list of 14 errors annotated for the video dataset]:

- Category: Initial incorrect position common to both jumps [6].
  - **Feet less than shoulder width apart**: If feet of the participant are closer to each other than half of the hip-width of the participant.
- Category: Initial incorrect position, during drop jump [12, 10].
  - **Jumped upward from box, rather than forward**: If jump is visually more upwards.
  - Asymmetric landing after jump: If one foot hits the force plate before the other.
- Category: Intermediate incorrect motion, during drop jump [3].
  - **Squat too low**: If the participant squats below their knee.
  - Heels touch force plate: If either heel touches the force plate for more than a couple seconds.
- Category: First or final landing on force plate, common to both jumps.
  - Knee collapse: If knees move visually more inwards when participant lands on force plate.
  - Both feet not on respective platforms: If either foot is not on respective force plate on landing.
  - Land off-balance: If participant lands off-balance or partially on the feet.
- Category: During jump, common to both jumps [5].
  - Off-balance: If the participant has an excess sideways body motion during the jump.

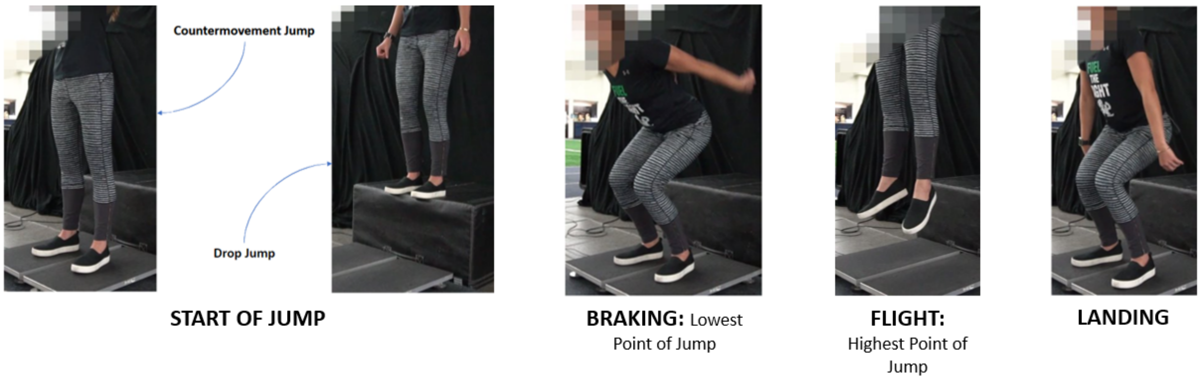

Figure 1. Important Events of an Evaluative Jump. The 'Start of Jump' is different for both jump types, while other events are similar.

  - Body twists, landing at different angle: If participant rotates about a vertical axis during jump and lands at different angle.

- Category: Final landing error, common to both jumps.

  - **Landing at different position from initial landing**: If either foot lands a slight distance away from their respective starting position.

  - **Excessive hip and knee flexion before returning to upright standing position**: If squat is lower than knee position.

  - **Take additional steps to maintain balance**: If balance lost on landing and participant takes additional steps to maintain balance.

  - Feet less than shoulder width apart: Similar to starting position.

1.4. Inter-observer agreement

Cohen's kappa is used as a measure for the agreement between two raters, and is useful in this case for evaluating if the rules defined for annotating the errors in the jumps can be used universally as a standard without changes in annotation outcomes. A second observer annotated an example set of 100 videos using the above set of rules, and with the help of an expert in health and exercise science. These annotations were then scored based on Figure 2, considering the chance agreement. Equations 1 refer to Cohen's Kappa calculations [Refer Table 1 for Kappa scores with respect to the error annotations].

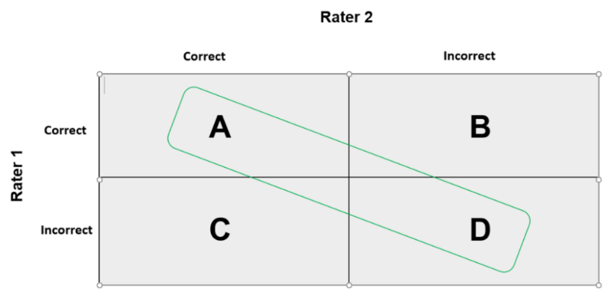

Figure 2. Understanding Cohen's Kappa. It is used as a measure for the agreement between two raters, and is useful in this case for evaluating if the rules defined for annotating the errors in the jumps can be used universally as a standard without changes in annotation outcomes.

$$\begin{cases} P_0 = Number\,in\,Agreement\,/\,Total \\ P_{correct} = (A+B/A+B+C+D) \\ \qquad\qquad \times(A+C/A+B+C+D) \\ P_{incorrect} = (C+D/A+B+C+D) \\ \qquad\qquad \times(B+D/A+B+C+D) \\ P_e = P_{correct} + P_{incorrect} \\ k(Kappa) = P_0 - P_e\,/\,1-P_e \end{cases} \tag{1}$$

1.5. Error intensity scores

While performing the evaluative jump motion, each mistake (error) that the participant might make would tend to have different levels of impact on their overall injury risk. We evaluate the intensity of injury risk with respect to each of the primary six errors, for a randomly sampled subset of 100 videos present in the dataset (error distribution in these videos is with respect to the error ratios present in the complete data). The following 5-point scale has been used to annotate the intensities of the errors (Scale range: 0.1 to 1.0):

Table 1. Errors in jump motion. 14 annotated errors (sub-categories) in the evaluative jumps, along with the number of samples in the dataset. 10 errors are annotated for both jump types, while 4 errors are specific to the drop jump. Errors in bold are the 6 primary errors used to train the classification models.

Errors (Overall categories)	Jump Type	Sr. No.	Errors (Sub-categories)	Inter-observer agreement (Subset of 100 video samples)		No. of samples in dataset
				Kappa (κ)	Samples in subset	
Start position	Both	1	Feet less than shoulder width apart	1.00	7	30
Initial position, after start	Drop jump	2	Jumped upward from box, rather than forward	1.00	2	22
		3	Asymmetric landing after jump	1.00	3	15
First landing on force plate	Drop jump	4	Squat too low	1.00	14	37
		5	Heels touch force plate	0.90	14	83
First or Final landing on force plate	Both	6	Knee collapse	0.72	19	64
		7	Both feet not on respective platforms	1.00	1	5
		8	Land off-balance	0.91	12	49
During jump	Both	9	Off-balance	0.75	8	79
		10	Body twists, landing at different angle	0.75	16	74
Final landing on force plate	Both	11	Landed at different position from initial landing	0.81	42	133
		12	Excessive hip and knee flexion before returning to upright standing position	0.88	16	78
		13	Take additional steps to maintain balance	0.73	7	94
		14	Feet less than shoulder width apart	1.00	1	3

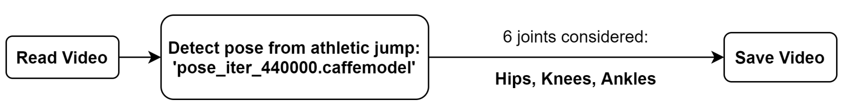

Figure 3. Steps to extract pose data from video.

- **0.10**: Lowest intensity, no expected harm due to jump errors.
- **0.25**: Lower intensity, jump errors are less harmful.
- **0.50**: Medium intensity, jump errors are harmful.
- **0.75**: Higher intensity, chances of injury if jump errors are not resolved.
- **1.00**: Highest intensity, chances of injury in immediate future.

The Fleiss Kappa [4] agreement scores are shown in Table 3, based on annotations provided by two raters.

Table 2. Error-wise model accuracies (%) of presented architectures. F-SW: Feet less than shoulder width apart; JU: Jumped upward from box, rather than forward; SL: Squat too low; LDP: Landed at different position from initial landing; EF: Excessive hip and knee flexion before returning to upright standing position; AS: Take additional steps to maintain balance. Number of instances for each error are provided in parentheses.

Experiment			Accuracy	F-SW (30)	JU (22)	SL (37)	LDP (133)	EF (78)	AS (94)
OpenPose	CF	Center	69.2	76.7	59.1	89.2	57.1	67.9	58.5
		Left	62.2	66.7	90.9	81.1	52.6	53.8	41.5
		Right	67.8	73.3	59.1	67.6	71.4	66.7	44.7
		Combined	72.4	76.7	90.9	91.9	77.4	76.9	65.9
	CKF	Center	64.6	76.7	68.2	83.8	46.6	65.4	44.7
		Left	63.4	86.7	95.5	91.9	48.1	51.3	40.4
		Right	67.0	76.7	63.6	64.9	72.2	71.8	46.8
		Combined	70.8	80.0	68.2	86.5	66.2	82.1	44.7
ResNet-18	+ LSTM	Center	62.9	76.7	86.4	81.1	71.4	87.2	69.1
		Left	66.1	63.3	50.0	45.9	60.9	69.2	54.3
		Right	62.9	73.3	81.8	64.9	71.4	82.1	68.1
		Combined	67.5	86.7	95.5	83.8	78.9	83.3	78.7
TSM		Center	64.5	96.7	81.8	81.8	72.9	87.2	72.3
		Left	61.1	80.0	63.6	48.6	50.4	53.8	52.1
		Right	62.5	80.0	68.2	45.9	59.4	64.1	62.8
		Combined	66.0	83.3	77.3	81.1	66.9	76.9	69.1

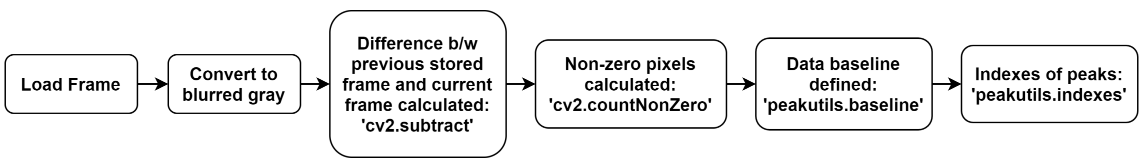

Figure 4. Steps to extract keyframes from video.

2. Method

2.1. Pose data preprocessing

The joint data is extracted using the following steps:

1. Video is read using the OpenCV function 'VideoCapture'.

2. Output video file is created, using required 'fps' and dimensions of frames in the video.

3. Pose is detected in a frame using the OpenPose [1] trained model 'pose_iter_440000.caffemodel'. A 0.3 threshold is used for confidence in the pose detection. This threshold ensures that most of the less confident pose estimations are eliminated, to provide a more stable temporal data on pose estimation for the participant jumps.

4. Six joints are considered when looking at the 0.3 threshold: Right Hip, Right Knee, Right Ankle, Left Hip, Left Knee, and Left Ankle. If atleast 3 of these joints are detected with a 0.3 confidence or more by OpenPose, then the frame is kept, else it is discarded.

5. This process repeats for all frames in a video, and the kept frames are stored in the output video file. The steps are summarized in Figure 3.

6. An extra step is performed for the second set of data, in which keyframes are selected as a subset from the confidently detected frames.

   (a) For getting keyframes, frames are first converted to blurred gray frames.

   (b) If the frame is the first of the video, it is saved in a variable.

   (c) Difference between the previous and the current blurred gray frames is calculated

   (d) Number of non-zero pixels in the image difference is calculated

Table 3. Error intensity annotations [Note: The error samples do not sum up to a 100, as there are instances where a video has more than one type of error].

ERRORS	SAMPLES IN SUBSET	KAPPA (κ)
FEET LESS THAN SHOULDER WIDTH APART	15	0.93
JUMPED UPWARD FROM BOX, RATHER THAN FORWARD	8	1.00
SQUAT TOO LOW	16	0.81
LANDED AT DIFFERENT POSITION FROM INITIAL LANDING	54	0.94
EXCESSIVE HIP AND KNEE FLEXION BEFORE RETURNING TO UPRIGHT STANDING POSITION	32	0.87
TAKE ADDITIONAL STEPS TO MAINTAIN BALANCE	43	0.90

(e) Using these non-zero pixels, the baseline of the data is defined

(f) Indices of the peaks are extracted: These are indexes of frames with the non-zero pixel count above a specified threshold.

(g) These indexes are then used to identify and store the keyframes. The steps are summarized in Figure 4.

3. Results and Analysis

3.1. Error-wise results

Table 2 provides error-wise accuracy scores of every architecture presented in the main paper. These results show their performance in correctly detecting the individual errors (the six primary errors are considered).

References

[1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*, 2018.

[2] Eamonn P Flanagan and Thomas M Comyns. The use of contact time and the reactive strength index to optimize fast stretch-shortening cycle training. *Strength & Conditioning Journal*, 30(5):32–38, 2008.

[3] Mitch Hauschildt. Landing mechanics: What, why, and when. `http://myweb.facstaff.wwu.edu/chalmers/PDFs/Landing%20mechanics.pdf`.

[4] Intelligent Natural Interaction Technology (INIT) Laboratory. How to calculate interrater reliability using fleiss' kappa, 2020. `https://init.cise.ufl.edu/2020/04/24/how-to-calculate-interrater-reliability-using-fleiss-kappa/`.

[5] Guillaume Laffaye, Phillip P Wagner, and Tom IL Tombleson. Countermovement jump height: Gender and sport-specific differences in the force-time variables. *The Journal of Strength & Conditioning Research*, 28(4):1096–1105, 2014.

[6] Nicholas P Linthorne. Analysis of standing vertical jumps using a force platform. *American Journal of Physics*, 69(11):1198–1204, 2001.

[7] John J McMahon, Timothy J Suchomel, Jason P Lake, and Paul Comfort. Understanding the key phases of the countermovement jump force-time curve. *Strength & Conditioning Journal*, 40(4):96–106, 2018.

[8] Frank R Noyes, Sue D Barber-Westin, Cassie Fleckenstein, Cathy Walsh, and John West. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. *The American journal of sports medicine*, 33(2):197–207, 2005.

[9] Jason S Pedley, Rhodri S Lloyd, Paul Read, Isabel S Moore, and Jon L Oliver. Drop jump: A technical model for scientific application. *Strength & conditioning journal*, 39(5):36–44, 2017.

[10] Randy J Schmitz, John C Cone, Amanda J Tritsch, Michele L Pye, Melissa M Montgomery, Robert A Henson, and Sandra J Shultz. Changes in drop-jump landing biomechanics during prolonged intermittent exercise. *Sports Health*, 6(2):128–135, 2014.

[11] Christopher J Sole. *Analysis of countermovement vertical jump force-time curve phase characteristics in athletes*. PhD thesis, East Tennessee State University, 2015.

[12] Robert Wood. "drop jump test.", September 2019. `https://www.topendsports.com/testing/tests/drop-jump.htm`.