28th ECSS Anniversary Congress, Paris/France, 4-7 July 2023

Detecting punches from multi-views videos data for performance analysis in elite boxing

VITEL, V., TALBI, I., SCHORTGEN, A., GOYALLON, T., SAULIERE, G., MOUSSA, I., REVERET, L., BREZEPHIN, L.

IRMES - INSEP - Université de Paris, INRIA - LJK - LOKI - ZENITH, Féderation Francaise de Boxe

INTRODUCTION:
The quantification of punches thrown and received by a boxer is a key performance indicator to assess boxing activity [1]. More than a scoring factor, hit quantification allows to evaluate the physical load endured, which is essential for injury prevention [2]. Manually counting punches is a tedious and time-consuming annotation workload. To overcome this challenge, we have developed an automatic method for punch recognition from videos.

METHODS:
To build our training dataset, we recorded 21 rounds of 3 min between elite boxers with 11 synchronized and calibrated cameras (Qualisys Miqus 60 fps, Full HD) distributed around the ring. For each point of view, we used a 2D pose estimator [3] to extract the body joint positions of both boxers at each frame in image coordinate system. The 2D skeletons are tracked and triangulated to avoid 2D issues (i.e. occlusions, depth estimation). In addition, 2508 punches were manually annotated at the exact frame of impact. We designed a temporal Convolutional
Neural Network [4] with Long Short Term Memory architecture to integrate spatial and temporal features tied to the action. The model outputs a hit probability for each frame throughout the video which allows to identify punch events with a 25 frames buffering sequence.

RESULTS:
We obtain an accuracy of 93.2% and an F1-score of 93.3% on our validation set. Our model's predictions showed a significant impact of parameters like the number of joints and the sequence length. For each boxer, our model creates a playlist containing video clips of all the detected punches. It can be used for a first analysis by coaches, and will help us to expand our dataset. Based on fist speed, we also identify which of the boxer's arms is used to deliver the punch which gives additional qualification of the punches (unknown beforehand).

CONCLUSION:
Our model quantifies punches thrown and received by a boxer with a good accuracy from multi-views feed. However, some detection errors persist due to specific gestures such as feints and combinations. Future work will focus on enriching our analysis by performing strike classification (i.e feints, hooks, jabs, uppercut), improving the model accuracy and deploying this approach inside the French Boxing training center. We aim to couple this automatic strike detector with more features such as impact zone or fist acceleration magnitude.

[1] S. E. Ashker, Technical and tactical aspects that differentiate winning and losing performances in boxing, International Journal of Performance Analysis in Sport (2011).
[2] J. M, et al, Boxing-related head injuries, Phys Sportsmed (2010).
[3] Z. Cao, et al, Realtime multi-person 2d pose estimation using part affinity fields, CVPR (2017).
[4] H. Duan, et al, Revisiting skeleton-based action recognition (2021).

Topic: **Statistics and Analyses**

Presentation   **Poster**

European Database of Sport Science (EDSS)

Supported by SporTools GmbH

34946