SportsVideo: A Multimedia Dataset for Sports Event and Position Detection in Table Tennis and Swimming

Aymeric Erades[1], Pierre-Etienne Martin[2], Romain Vuillemot[1,*], Boris Mansencal[3], Renaud Peteri[3], Julien Morlier[3], Stefan Duffner[4] and Jenny Benois-Pineau[3]

[1]*Ecole Centrale de Lyon, LIRIS, France*
[2]*CCP Department, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany*
[3]*University of Bordeaux, Labri, France*
[4]*INSA Lyon, LIRIS, France*

Abstract
Positions and actions detection/classification are one of the main challenges in visual content analysis and mining. Sports video offer such challenges due to the variety of scenes and actions they contain. Sports also provide a wide range of analysis, related to athletes' performances and tactics. We propose a series of 6 sports-related tasks, divided each into 2 sub-tasks for two sports, table tennis and swimming. Those tasks are a follow-up from the Sport Task and SwimTrack task from the 2022 MediaEval Benchmark.

1. Introduction

We present SportsVideo, a series of six sports-related multimedia tasks, divided in sub-tasks for table tennis and swimming. The dataset is a merge of the Sport Task (Table Tennis) [1] with SwimTrack [2] (Swimming) task from the Medieval 2022 edition [3] into a single benchmark dataset. The main motivation is to provide a more complete and challenging benchmark for video analysis with complex scenes and actions. The first four tasks are related to image and video analysis, the fifth to sound analysis and the last one to textual information extraction.

By combining two different sports –table tennis and swimming– we aim to encourage an approach that generalizes beyond a single sport type since the two sports are very different in terms of the type of video, the type of events and the type of analysis required. Table tennis is a fast-paced sport with a lot of action and a small field of view. Swimming is a slower sport with a large field of view and a lot of occlusions. We expect participants to develop approaches that can generalize to both, but also to other sports. Those tasks have been identified and designed to be as independent as possible so that participants can choose to participate in one or more tasks. But if combined, they can provide a more complete analysis of sports videos. Participants are encouraged to release their code publicly with their submission. This year, similarly to the Sport Task 2022 edition [4], a baseline for both subtasks 2.1 and 3.1 is shared publicly[1] [5]. Background on sports is provided in the following two PhD thesis in swimming [6] and in table tennis in [7].

MediaEval'23: Multimedia Evaluation Workshop, February 1–2, 2024, Amsterdam, The Netherlands and Online
*Corresponding author.
pierre_etienne_martin@eva.mpg.de (P. Martin); romain.vuillemot@ec-lyon.fr (R. Vuillemot)
www.eva.mpg.de/ccp/staff/pierre-etienne-martin (P. Martin)
0000-0002-9593-4580 (P. Martin)

[1]https://github.com/ccp-eva/SportTaskME23

2. Tasks Description

Task 1 - positions detections

The main information in sports is related to positions of players in videos featuring different numbers of lanes in a swimming pool and sides around a table tennis (seen from various angles). They need to provide bounding boxes for identified players, and their results are evaluated using Average Precision (AP) at an IoU ratio of 0.25, counting true positives and negatives across the dataset.
Subtask 1.1 (table tennis) – To detect 2 or 4 players (depending if single or double) and track them during the video especially during double games where players have a lot of overlaps, from videos recorded from various angles (e.g., side, corder).
Subtask 1.2 (swimming) – To detect up to 8 swimmers in the pool from static videos (recorded from the side of the pool). A baseline is provided in [8].

Task 2 - events detection

Another key information in sports video is related to events, in particular strokes, which are key for performance evaluation. In swimming, stroke events indicate the pace of swimming for Freestyle, Backstroke, and Butterfly are triggered when the right hand enters the water, while for Breaststroke, when the head reaches its highest point.
Subtask 2.1 (table tennis) – To detect when a player is performing a stroke (i.e. a ball hit with the racket) using close-up videos. The goal is to detect the exact frame when the ball is hit by the racket. Videos are provided with the ball annotated. Evaluation is based on the F1-score, which measures the harmonic mean of precision and recall.
Subtask 2.2 (swimming) – To detect each time a swimmer is achieving a repeated motion for each swimming style (for freestyle, backstroke, and butterfly stroke once the swimmer's right hand enters the water; for breaststroke once the head is at its highest point). Swimmers' strokes are identified after the underwater phase and until the swimmer finishes the race, excluding underwater phases for races longer than 50m. Video clips of cropped swimmers are provided, and evaluation is based on Off-By-One Accuracy, which measures the proportion of correctly estimated stroke counts within a tolerated error of one stroke in the dataset. A baseline is provided in [9].

Task 3 - events classification

The goal of this task is to classify the type of stroke performed by a player in table tennis and swimming. Participants need to categorize a collection of shortened table tennis stroke videos, each containing either a single stroke or no stroke. There are 20 potential stroke categories and an extra category for non-strokes. Two sets with annotations are given: a training set with 807 videos and a validation set with 230 videos. The challenge is to classify a non-annotated test set comprising 118 videos, with the assurance that the trimmed videos in each set are derived from the same untrimmed videos but captured at different time instances without temporal overlap.
Subtask 3.1 (table tennis) – To classify different strokes in table tennis from trimmed videos in which only one stroke is present. There are 3 different categories of strokes, services, forehand and backhand. For services we have 6 different classes. For forehand and backhand we have 5 classes. For a total of 16 classes and one non-stroke class.
Subtask 3.2 (swimming) – To classify different swimming styles (Freestyle, Backstroke, Breaststroke, Butterfly).

Task 4 - field/table registration

Sports videos in general, and the one we provided, are usually recorded from the side. This tasks asks participants to find the absolute homography matrix for each frame in the dataset. The precision of this projection is evaluated using Intersection over Union (IoU), with two metrics: IoU for the visible pool parts and IoU for the entire pool, including parts outside the camera's field of view.
Subtask 4.1 (table tennis) – To detect the table position for a given video frame of a whole table tennis. The dataset contains 54 annotated images with homography matrix from TV broadcasts of table tennis matches.
Subtask 4.2 (swimming) – To detect pool position for a given video frame. The dataset is provided with 500 annotated images with homography matrix from [10]

Task 5 - sound detection

Sports are highly multi-modal events. Sound is an important modality that can be used to detect events. They can either be used as official cues (e.g., buzzer sound in swimming) or as additional cues (e.g., ball bounce in table tennis). In this task, participants are asked to detect sound events in table tennis and swimming videos. In table tennis, the ball bounces on the table at each stroke.
Subtask 5.1 (table tennis) – Ball hits indicate the pace of the game. The goal is to detect the exact frame when the ball bounces on the table. Videos are provided with the ball annotated and evaluation is based on the F1-score.
Subtask 5.2 (swimming) – A buzzer sound (preceded by an "on your mark") signals the start. Participants are asked to find the time of this sound within audio files extracted from live videos. These files may or may not include the buzzer sound, which can occur at various points during the recording. This task is challenging because the sound might be recorded from a considerable distance and could be accompanied by significant background noise.

Task 6 - score and results extraction

In most sports, the outcome is presented on a scoreboard, featuring race time for each swimmer (and possibly extra data like reaction time) or the current score of the game. These scoreboards are typically either physical LCD screens on the wall or close to the referee. Digital versions shown on TV broadcasts.
Subtask 6.1 (table tennis) – To recognise the score of the match. In table tennis, the score of a match can be embedded in the broadcast video or it can be shown by referees with scoreboards. When score is embedded in stream video, names of players are also displayed.
Subtask 6.2 (swimming) – The task here involves extracting swimmers' names, lane numbers, and race results (times) from screenshots of these scoreboards. The images and scoreboard coordinates will be provided, with the localization aspect already addressed. During swimmer competitions, after each race, results are displayed on digital boards. The goal is to recognise characters of these boards to obtain the results of races.

3. Acknowledgement

We thank the people involved in the previous versions of this challenge. This project was partially funded by the ANR NePTUNE, grant number ANR-19-STHP-0004 and the FFTT partnership convention.

References

[1] P. Martin, J. Calandre, B. Mansencal, J. Benois-Pineau, R. Péteri, L. Mascarilla, J. Morlier, Sport task: Fine grained action detection and classification of table tennis strokes from videos for mediaeval 2022, in: [3], 2022. URL: https://ceur-ws.org/Vol-3583/paper26.pdf.

[2] N. Jacquelin, T. Jaunet, R. Vuillemot, S. Duffner, SwimTrack: Swimmers and Stroke Rate Detection in Elite Race Videos, 2023. URL: https://hal.science/hal-03936053. doi:10.1145/nnnnnnn.nnnnnnn.

[3] S. Hicks, A. G. S. de Herrera, J. Langguth, A. Lommatzsch, S. Andreadis, M. Dao, P. Martin, A. Hürriyetoglu, V. Thambawita, T. S. Nordmo, R. Vuillemot, M. A. Larson (Eds.), Working Notes Proceedings of the MediaEval 2022 Workshop, Bergen, Norway and Online, 12-13 January 2023, volume 3583 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3583.

[4] P. Martin, Baseline method for the sport task of mediaeval 2022 with 3d cnns using attention mechanism, in: [3], 2022. URL: https://ceur-ws.org/Vol-3583/paper19.pdf.

[5] P. Martin, Baseline method for the sport task of mediaeval 2023 3d cnns using attention mechanisms for table tennis stoke detection and classification., in: Working Notes Proceedings of the MediaEval 2023 Workshop, Amsterdam, The Netherlands and Online, 1-2 February 2024, CEUR Workshop Proceedings, CEUR-WS.org, 2023.

[6] N. Jacquelin, Automatic Analysis of Elite Swimming Race Videos, These de doctorat, Ecully, Ecole centrale de Lyon, 2022. URL: https://www.theses.fr/2022ECDL0017.

[7] P.-E. Martin, Fine-grained action detection and classification from videos with spatio-temporal convolutional neural networks : Application to Table Tennis., Theses, Université de Bordeaux, 2020. URL: https://theses.hal.science/tel-03128769.

[8] N. Jacquelin, R. Vuillemot, S. Duffner, Detecting Swimmers in Unconstrained Videos with Few Training Data, 8th Workshop on Machine Learning and Data Mining for Sports Analytics (2021).

[9] N. Jacquelin, R. Vuillemot, S. Duffner, Periodicity Counting in Videos with Unsupervised Learning of Cyclic Embeddings, Pattern Recognition Letters (2022). URL: https://hal.archives-ouvertes.fr/hal-03738161. doi:10.1016/j.patrec.2022.07.013.

[10] N. Jacquelin, S. Duffner, R. Vuillemot, Efficient One-Shot Sports Field Image Registration With Arbitrary Keypoint Segmentation, in: IEEE International Conference on Image Processing, Bordeaux, France, 2022. URL: https://hal.archives-ouvertes.fr/hal-03738153.