SwimTrack: Swimmers and Stroke Rate Detection in Elite Race Videos

Nicolas Jacquelin[1,2], Théo Jaunet[1,2], Romain Vuillemot[1,2,*] and Stefan Duffner[1,3]

[1]*LIRIS CNRS UMR 5205*
[2]*École Centrale de Lyon, France*
[3]*INSA Lyon, France*

Abstract

We present SwimTrack, a series of 5 multimedia tasks related to swimming video analysis from elite competition live recordings. These tasks are related to video, image, and audio analysis which may be achieved independently. But when solved altogether, they form a grand challenge to provide sport federations and coaches with novel methods to asses and enhance swimmers' performance, in particular related to stroke rate and length analysis. We share a unique collection of video footage that contains all swimming race types, recorded from a spectator point of view with variations such as lighting reflections, background clutter, noise from the motion of waves, and different point of views on swimmers. SwimTrack is the first challenge of this kind for a total of 4 swimming elite competitions. We sought to include a larger and even more diverse set of videos as well as additional mini-challenges once more recordings will be available in a next version.

1. Introduction

Swimming is one of the most ancient, yet popular Olympic disciplines as it holds the second highest potential of gold medals (37) after athletics (48). It has a long tradition of being analyzed quantitatively (e. g., race time, lap time, rankings) due to official time recording devices. There is however little information at a more detailed level, i. e., within laps or on the swimmers' speed and real-time motion, except for manually annotated datasets. Recent efforts in deep learning (e. g., [1, 2]) have proposed approaches that pave the way for automated fine-grained data extractions, but a very high level of accuracy and robustness is yet to be reached to rely on their application to any swimming pool or camera position.

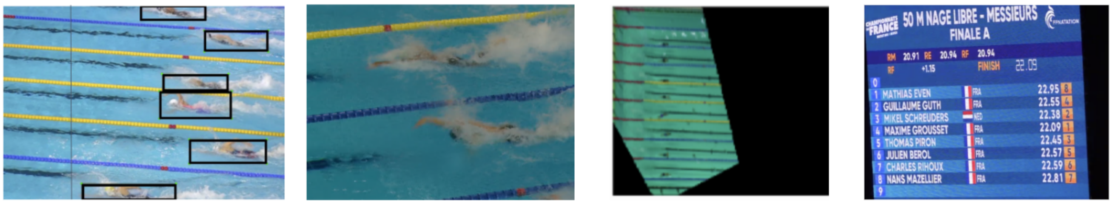
The goal of the SwimTrack challenge is to push the envelope of systems that accurately track swimmers' motion in a reliable way during elite competitions. Current state of the art in multi-object tracking is limited by the unusual nature of a swimmer's motion and large noise generated by the water. This first version of the challenge is divided into 5 independent tasks. Each of them contains its own set of input data, output format, and an evaluation metric. Participants are required to follow this format to get feedback and make improvements on their contribution. We will proceed with a classical evaluation protocol: expected metrics for each task will be tested with a first dataset *VALIDATION* which contains both input and ground truth. A second dataset *TEST* will be kept private by the organizers, to prevent participants from exploiting solutions to reach better results (e. g., overfitting a model for each individual task). Thus, participants are free to use the *VALIDATION* in any way they want to build their solution,

MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online

*Corresponding author.
†These authors contributed equally.
nicolas.jacquelin@ec-lyon.fr (N. Jacquelin); theo.jaunet@ec-lyon.fr (T. Jaunet); romain.vuillemot@ec-lyon.fr (R. Vuillemot); stefan.duffner@insa-lyon.fr (S. Duffner)

CEUR Workshop Proceedings (CEUR-WS.org)

but they need to make sure it is not too specific to it. They will however have a restricted number of *TEST* tokens to get the score on their solution during a time-limited testing period.



task 1 (position) task 2 (cycles) task 3 (registration) task 4 (ocr)

Figure 1: Images samples from the 4 visual tasks (excluding our 5th task which relies on the analysis of sounds). From left to right, **1**, swimmers' position detection, **2** stroke rate detection, **3** camera registration, and **4** character recognition of the score board. Task **5** is non-visual as related to buzzer sound detection of the start of the race.

2. Video Collection

We will provide swimming videos recorded during elite competitions from a fixed spot on the stands for task 1, numeric zoom using crops for task 2 and pan/zoom for task 3. Videos cover all the 4 swimming styles (Freestyle, Backstroke, Breaststroke, Butterfly, Medley), both genders (female, male) and principal race lengths (50m, 100m, 200m, 400m) for 50m-long swimming pools. They cover all the swimming phases (e. g., standing, diving, underwater, return and finish). The camera view parameters vary from static wide angle to zoomed + moving using various camera types (GoPro 8, Blackmagic Pocket 6K and Panasonic HC-V750). Resolutions range from HD to 4K with variable frame rates across the recordings (between 25fps and up to 50fps). Despite those differences, the provided videos share the same MP4 format resulting from the same compression algorithm. Dataset samples and utils will be made available on https://github.com/centralelyon/swimtrack and the full dataset is provided upon request. This website will also collect feedback using GitHub issues to answer the most frequently asked questions and will provide code to parse and process the provided datasets when needed. As videos have been recorded during public event there will be no privacy issue.

3. Task Description

We now describe the 5 tasks that can be achieved independently and which image samples are displayed on Figure 1.

Task 1: Swimmer Position Detection

In this task, participants are requested to estimate swimmers' positions in a swimming pool. The input are videos with a number of occupied swimming lanes by swimmers, ranging from one to ten. Answers must be provided as bounding boxes on the video for each identified swimmer. As for general object detection algorithms in computer vision, the results will be evaluated by calculating the overlap between the participants' answers and the ground truth, namely the Average Precision (AP) 25, i. e., if a true box is overlapped by an estimated box with an Intersection over Union (IoU) ratio greater than 0.25, it is counted as positive, if not, as negative. The AP25 is the ratio: #Positives / (#Positives + #Negatives) across the whole dataset. A baseline result is also provided from the method in [2] along with manual annotations.

Task 2: Stroke Rate Detection

Here, participants have to identify swimming strokes events, an important information to further calculate the stroke rate during a race. For Freestyle, Backstroke, and Butterfly a stroke is triggered once the swimmer right hand enters the water. For Breaststroke a stroke is triggered once the head is at its highest point. The strokes will be identified once the underwater phase has ended, and until the swimmer has not yet finished its race (also excluding underwater phases for races longer than 50m). For this reason, we will provide video clips of cropped swimmers excluding the underwater phases following dives and returns. Evaluation will be performed by measuring the commonly used Off-By-One Accuracy (OBOA) [3]: it counts the proportion of videos in the dataset with a correctly estimated number of strokes, up to a tolerated error of one stroke. For instance, with 3 videos from the dataset, if the first prediction is the correct number of strokes (OK), the second has a difference of 2 strokes (not OK) and the last has an error of 1 strike (OK), then $OBOA = 2/3 = 0.66$. Video races with stroke rate will be given along with ground truth [4].

Task 3: Camera Registration

The races are shot from the side of the pool or the stands, thus, due to the perspective and geometrical projection on the image, its shape is not rectangular and only partially visible. To compensate this effect, one can use a homography projection to create a virtual top-view of the pool. In this tasks, participants have to find the (absolute) homography matrix corresponding to each frame provided in the dataset. The precision of such projection is measured using the IoU between the ground truth top-view and the estimated one (see Fig. 1, task 3). We will use two metrics: IoU^{part} which compares only the pool's visible parts of the top-views, and IoU^{whole} which uses the whole pool, i. e., even the parts that are outside the camera's field of view. The average and median of these metrics are used. The dataset, called RegiSwim^{500}, can be found at https://github.com/njacquelin/sports_field_registration with 500 annotated images with homography matrix from [5]

Task 4: Characters Recognition of Score Boards

The result of each race is displayed on a scoreboard displaying the race time, swimmer names, and sometimes additional information (e. g., reaction time). Such scoreboards are usually displayed on a physical LCD screen located on the swimming pool wall, or a digital version is shown in the TV broadcast. In this task, the objective is to extract swimmers' name, lane numbers, and their race result (time) from screenshots of such boards. Images will be provided together with the images coordinates of the scoreboard (i. e., the localization has already been done in this task). Evaluation will be conducted as follows:

- the swimmer names precision will be calculated using the edit distance between the prediction and the ground truth;
- race results are compared using the average absolute time difference between the prediction and the ground truth (MAE).

Task 5: Sound detection

Every swimming race starts with a buzzer sound (preceded by the iconic *on your mark*). Participants have to estimate when such a sound occurs in audio files extracted from live videos. The files may or may not contain a buzzer sound, which may occur at any time during a recording.

This task is far from trivial as the sound may be captured from a rather long distance and contain a large amount of background noise. Evaluation will be based on a precision-recall curve obtained by measuring the correct/missed detections and varying the tolerated absolute time difference between the predicted moments and the ground truth. The dataset contains sounds from TV recordings, but also in swimming pools from different locations.

4. Evaluation Protocol

Participants' proposal will be evaluated when submissions of solutions to our website will be permitted. This website will dynamically calculate the score for the TEST dataset of each task. If all tasks have been addressed, a general "grand challenge" score will be calculated. As stated in the introduction, we will however limit the number of times participant can submit solutions.

5. Quest for Insight

Our experience of closely working with sport performances teams of a national federation and coaches led us to identify the following questions they sought to answer:

- is stroke rate constant within and between laps?
- to increase their speed, do swimmers increase their stroke rate or their stroke length?
- how swimmers change their swimming strategy to cope with various constraints (e. g., fatigue, tight competition, bad start)?
- are there typical stroke rate profiles of swimmers? e. g., can they be categorized based on their stroke rate?
- how does a swimmer's stroke rate and length profile evolve through his/her career ?

6. Future Versions of the Challenge

In the future, we plan to augment the datasets in volume of races to provide even more diverse conditions. We plan to cover longer videos of 800m and 1.5k long races. Many competitions occur on 25m-long swimming pools (e. g., International Swimming League, pending copyright permissions). Finally we plan to include swimming race of athletes with disabilities to help them get prepared for the upcoming International competitions and Paralympic Games.

7. Acknowledgement

We thank Renaud Jester for his contribution to the the first version of the dataset. We also thank Clément Yvernes, Pierre Gallou-Guyot and Vinoth Dilipkumar for augmenting the dataset with additional data. We thank the performance center of the French Swimming Federation for videos. This project was partially funded by the ANR NePTUNE, grant number ANR-19-STHP-0004 and CNRS Prim'80 program.

References

[1] T. Woinoski, A. Harell, I. Bajić, Towards Automated Swimming Analytics Using Deep Neural Networks, ArXiv (2020).

[2] 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).

[3] T. F. H. Runia, C. G. M. Snoek, A. W. M. Smeulders, Real-world repetition estimation by div, grad and curl, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 9009–9017.

[4] 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.

[5] 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.