DEVELOPMENT OF TOOL FOR ANALYSIS OF SWIMMING USING POSE ESTIMATION ALGORITHM

Itay Coifman[1], May Hakim[1], Gera Weiss[2], Raziel Riemer[1*]
[1]Industrial Engineering and Management Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2]Computer Science Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel,
*Corresponding author's email: rriemer@bgu.ac.il

Introduction: Swimming poses unique challenges in biomechanical research due to the complexity of analyzing human movement, including kinematics, kinetics, and neuromuscular aspects, in an aquatic environment [1,2]. Past studies have developed underwater computer vision models based on deep learning for human pose estimation [3,4] focusing on predicting joint positions. Yet, few studies have used such models for biomechanical feature extraction [5] or classification [6], and none has attempted to provide a comprehensive solution for general biomechanical analysis in swimming. This study introduces an end-to-end system for analyzing and tracking swimming biomechanics using a single uncalibrated underwater camera (e.g., GoPro®, iPhone). The aim was to evaluate the accuracy of features extracted by the system, including stroke rate [7], elbow angle at push time, and shoulder rotations [8], which are commonly used in swimming studies. The primary focus was not to replace motion capture systems but rather to explore how simple 2D data can deliver insights on swimming.

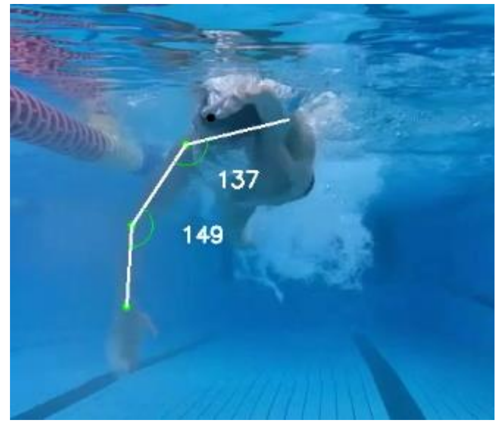

Figure 1: Example of underwater frame annotated by the human pose estimation model algorithm; annotated angles represent the joint angle at t_{push}, the point in the front crawl cycle when the hand starts moving backward (relative to the body).

Methods: We tested several human pose estimation algorithms, but none could identify the human joints underwater. Therefore, we trained the YOLO-V8 Pose by Ultralytics on transverse and sagittal videos of swimmers captured with various cameras and manually labeled (**Fig. 1)**. To evaluate the system's capabilities, we created a test set of spatial-temporal features typically used in swimming analyses. The data comprised 16 transverse and 12 sagittal front crawl videos not previously seen by the system. The kinematic data of this test set were manually labeled using Kinovea®. Keyframes for the start of the push phase (t_{push}) (**Fig. 1**) and hand entry (t_{entry}) were identified. Stroke rate was calculated as the average frequency between consecutive right-hand entries. Additionally, the 2D angles of the elbows in the transverse plane were measured from images at each t_{push} of the right hand. Shoulder rotation in the transverse plane was measured at t_{push} of both the right and left hands (i.e. two measurements for each cycle). We calculated the average error over all cycles for each video (average of two cycles per video). The reported errors correspond to the mean error, with each video considered as one sample.

Results & Discussion: As **Table 1** shows, the temporal features exhibit very low errors, which indicates that the system can predict the timing of specific events during the swim cycle with high accuracy. In cases where the angle errors are less precise, the accuracy of 2D angles was significantly influenced by minor errors introduced by the pose estimation model. One notable limitation is that angle comparisons were made against human annotations rather than a gold standard comparison method, such as a motion capture system. These annotations are prone to inaccuracies, particularly for small angle differences. We anticipate that more precise data will be available in the future to enable a more accurate comparison of our system.
The mean absolute percentage error (MAPE) of the shoulder rotation is relatively large (16%), whereas the root-mean-square error (RMSE) is relatively small (approximately 7 degrees), which is attributed to the relatively smaller body rotations measured during a swimming cycle. We believe that more training data will significantly improve the pose estimation accuracy driving the errors towards zero.

Feature name	Average GT	RMSE	MAPE
Stroke rate (Hz), transverse	0.47	0.11	1.74%
Stroke rate (Hz), sagittal	0.47	0.001	0.19%
Elbow angle (Deg)	122.84	8.93	7.17%
Shoulder rotation (Deg)	22.01	6.96	16.11%

Table 1: Average error metrics, with each video constituting a single sample; average Ground Truth (GT) represents the average actual value annotated in Kinovea®.

Significance: To our knowledge, this study represents the first application of monocular computer vision algorithms to analyze swimming biomechanics. Its significance lies in its demonstration that essential biomechanical features can be accurately extracted even in challenging video conditions (e.g., bubbles, color attenuation). To extract these features accurately, it is important to understand the limitations of the analysis and use features that are accurately represented in the 2D image projection. We believe that our approach can support studies with larger sample sizes.

Acknowledgments: This study was partially supported by the Israel Innovation Authority and the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative of Ben-Gurion University of the Negev.

References: [1] Zecha et al. (2018), *CVPR Workshops*; [2] Fani et al. (2018), *ICIP*; [3] Giulietti et al. (2023), Sensors 23(4); [4] Zecha et al. (2019), *CVPR workshops*; [5] Zecha et al. (2012), *Multimedia on Mobile Devices (8304)*; [6] Einfalt et al. (2018), *IEEE WACV*; [7] Morais et al. (2022), *Front Physiol(13)*; [8] Vila Dieguez, Oscar, and John M. Barden (2022), *Sports Biomech 21(10);*