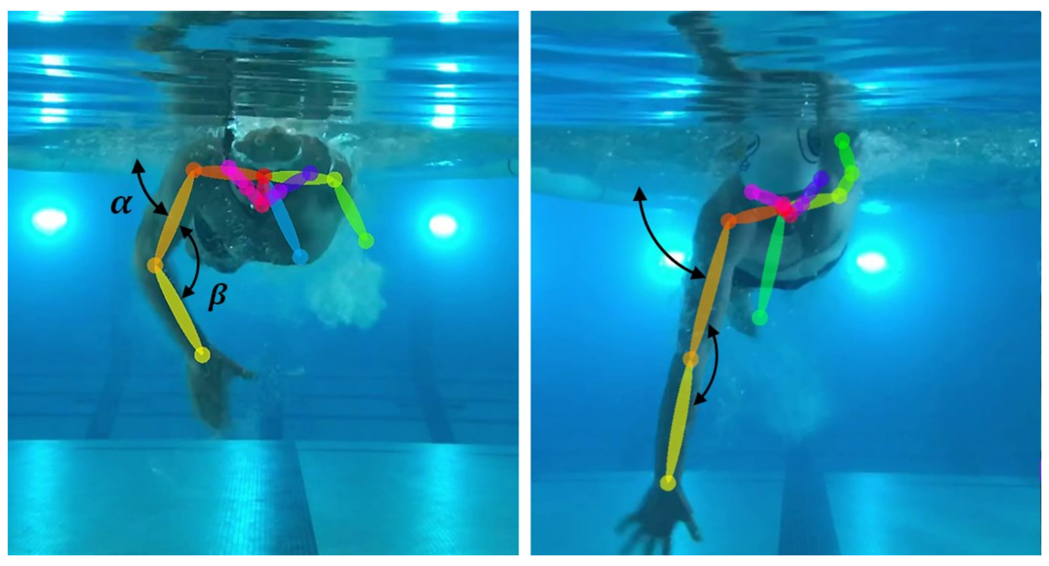
SWIM STROKE ANALYTIC: FRONT CRAWL PULLING POSE CLASSIFICATION

Hossein Fani[1], Amin Mirlohi[1], Hawre Hosseini[2], Rainer Herpers[3]

[1]Faculty of Computer Science, University of New Brunswick, NB, Canada
[2]Department of Electrical & Computer Engineering, Ryerson University, ON, Canada
[3]Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany

Problem Definition

In this work, we automatically distinguish the efficient high elbow pose from dropping one in pulling phase of front crawl stroke in front view amateurly recorded videos. This task is challenging due to the aquatic environment and missing depth information.



In sports, athletes attain their maximum potential by reviewing their own workouts in practice. In this respect, we focus our study on front crawl swimming style and predict the efficiency of the arm stroke under water in order to provide immediate feedback on poses' efficiency to the coaches as well as athletes themselves.

Approach

OpenPose represents the first real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints (in total 135 keypoints) on single images. The approach uses a nonparametric representation, which we refer to as Part Affinity Fields (PAFs), to learn to associate body parts with individuals in the image. The architecture encodes global context, allowing a greedy bottom-up parsing step that maintain high accuracy while achieving realtime performance, irrespective of the number of people in the image.

We detect the swimmer body parts using the state-of-theart pose detection method OpenPose [7]. The choice of Open-Pose is motivated for its functionality on image or video taken by webcam and IP cameras. This provides huge benefit in comparison to the skeletal tracking capability of Microsoft Kinect or the likes which depend on depth information, i.e. three dimensional camera. We then measure the angles between upper arm and forearm, and upper arm and the water surface. Given the arm joints position and the respective angles, we train a classifier on our manually labeled dataset of swimmers with different levels of expertise and physiological properties.

To learn the pulling classifier, we perform a supervised learning method on a set of frames and their respective pulling labels. Our proposed approach can be described in two steps: i) pose extraction, and ii) model training. To learn the pulling classifier, we map each frame to a feature space in which the arms' poses and angles in the frame are the features. The feature vector is a concatenation of two types of arm's skeletal information as follows:

- Joints: the 2-dimensional position of the right and left shoulders, elbows, and wrists
- Joint angles: the angle between upper arms and the water surface, namely α, and the angle between upper arms and forearms, namely β.

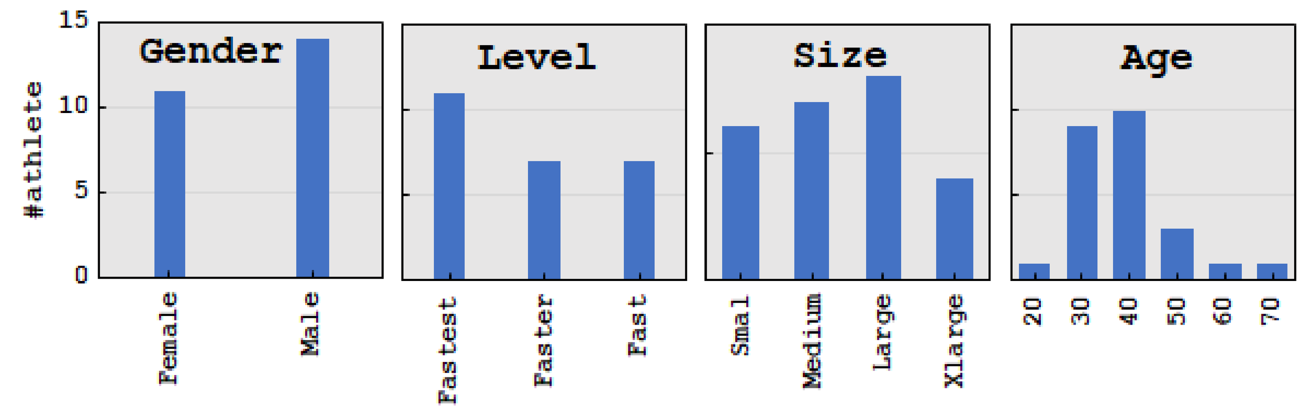
To extract the arm's joints, we rely on the already trained model of OpenPose to detect the swimmers' arm joints under water. We learn the pulling classifier function through the classification methods, namely random forest and svm over feature vectors. Random forest is an ensemble method which averages the predictions of several decision trees as base classifiers yielding an overall better model over a single decision tree. More, random forest due to its base classifier is inherently able to learn multi-class classifications such as our pulling classifier. The number of trees is set to 100 and the criterion to measure the quality of a node split is Gini impurity for the information gain in our random forest. As an alternative baseline, we also use svm with oneversus- rest (ovr) multi-class strategy and linear kernel.

Experimental Setup

We used iPhone 7 and LG G6 both with 1,920*1,080*30fps to record the swimmers front crawl strokes under water. We fixed the camera 10 centimeters under water on the one end of the swimming pool wall and parallel to the water surface. In order to identify the body parts and poses, we extracted the video's frames and applied OpenPose library with its default settings, i.e., COCO model to identify 17 body parts with neural net resolution 656*368. We build and run OpenPose library on Intel Core i7-3770 quad-core processor with 16GB DDR3 of system memory and NVIDIA 1070 graphic card with 1,920 GPU cores and 8GB frame buffer. We used scikitlearn3 to train and evaluate our baseline classifiers.

Data Acquisition

Our experiments include 25 swimmers of masters' swim club at Ryerson University4. The swimmers are from different levels of expertise. Fast (beginner), faster (novice), and fastest (professional). Also, they are sampled from a wide range of swimmers with different genders, ages, and body sizes.



As a result, the effect of physiological characteristics has been tried to be relieved. We record swimmers' front crawl pulling strokes under water from the front view as they approach, from one side of the pool, to the wall where our camera is installed (one lap swim of 25 yd = 22.86 m). We then extract frames and filter out those in which no arm joint has been detected by OpenPose. This way we filter out the starting frames where no trace of swimmer has been detected either by camera or OpenPose. This leads us from a dataset of 15,384 frames to 2,633 frames.

Evaluation and Results

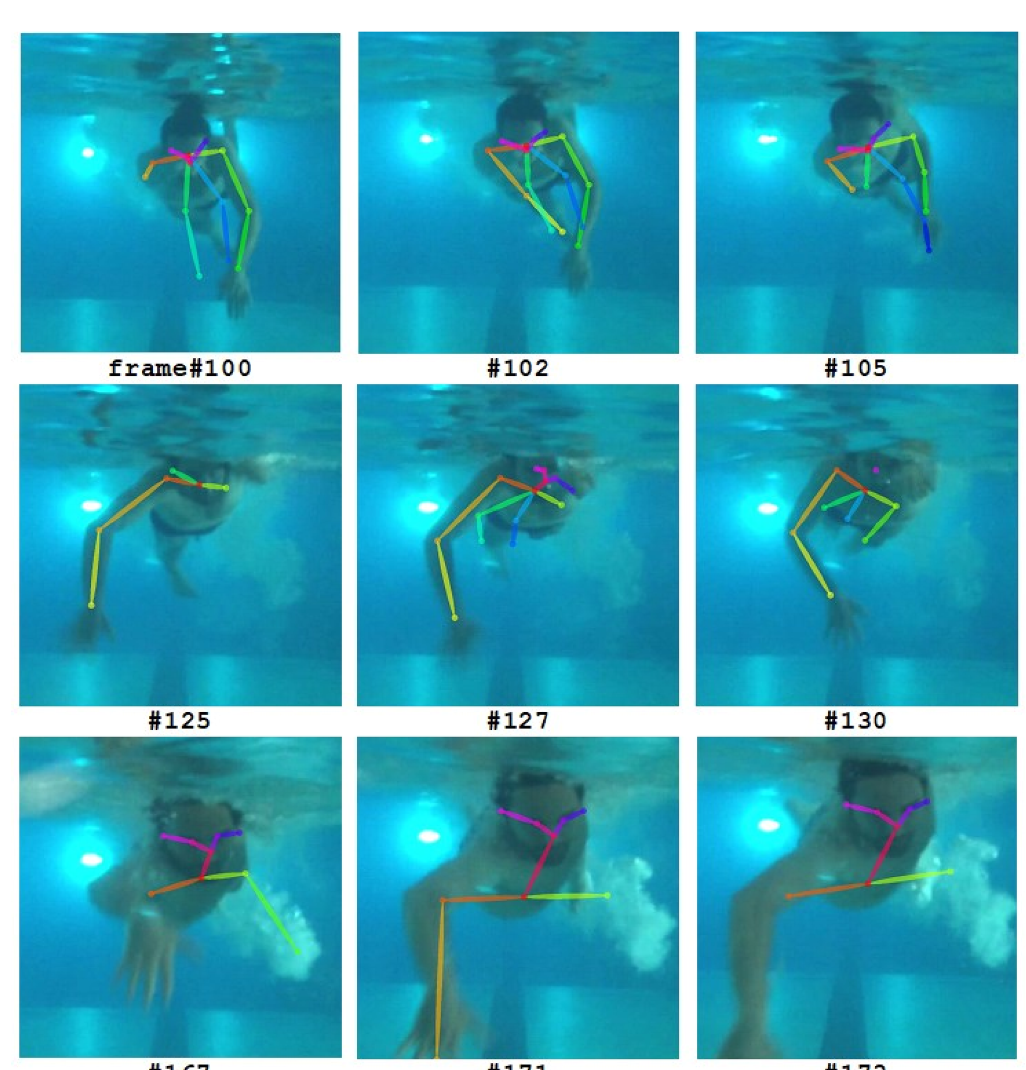
model	feature set	precision	recall	f-measure	accuracy
random forest	joints+angles	**0.652**	**0.666**	**0.649**	**0.666**
	joints	0.643	0.656	0.64	0.656
	angles	0.546	0.557	0.547	0.557
svm	joints+angles	**0.482**	**0.498**	**0.442**	**0.498**
	joints	0.457	0.439	0.383	0.439
	angles	0.422	0.437	0.396	0.437

Table: The performance of our baselines.

We evaluate our classifiers, random forest and svm, on our dataset with three different feature subsets each of which includes only i) joints, ii) joint angles (and), and iii) both joints and angles. Due to label distribution imbalance, we did stratified 10-fold cross-validation and report the performance of each baseline by weighted average of all three classes for precision, recall, f1-measure, as well as accuracy in Table.

Discussion and Future Direction

The goal of this study is to give feedback to the swimmer about efficiency of her pulling under water. We introduce the concept of swimmer pulling map which is able to give feedback on the swimmer's pulling efficiency over the course of different frames. As seen in the frames in Figure, while the swimmer is doing dropping elbow on his left arm, he is doing high elbow on the right arm. Presumably, the swimmer is right handed and has some frailty in his left arm and needs to practice more on his left arm pulling. As seen, although there are misclassifications



Two possible future directions to this work are: (1) the used pose detection library, OpenPose, is not perfect and shows false detection as in Figure (last row). This is due to the fact that we use the pre trained model of OpenPose in pose estimation which is not specifically trained for underwater environment. Moreover, we have not included any image prepossessing step on the video frames in our approach. These were intentional since we attempted to show the performance of our work with bare minimum configuration. An improvement to our work would be re-training the OpenPose's pose estimation model on the datasets of preprocessed frames where the positions of the joints are labeled as well already. (2) At a higher application level, we aim to extend our work to breaststroke and butterfly swim styles as the pulling pose analysis is very similar to front crawl.