Analysis of Waves and Human Interaction Using Beach Webcams

Harel Mendelman Tomer Masssas Yair Moshe
Signal and Image Processing Laboratory (SIPL),
Andrew and Erna Viterbi Faculty of Electrical & Computer Engineering
Technion – Israel Institute of Technology, Haifa, Israel, https://sipl.eelabs.technion.ac.il

***Abstract*—Automatic monitoring of beach video streams is important for improving safety, environmental monitoring, research, and education related to beach activities. This paper introduces a novel approach for monitoring water bodies by analyzing beach video streams. In contrast to earlier works, in the proposed approach we analyze not only the behavior of water bodies, or of humans on beach videos, but also the interactions between them. By integrating human activity and water behavior analysis, the approach provides new insights that are unattainable by analyzing each separately. To accomplish this objective, deep neural networks are utilized to analyze video streams from existing beach webcams. The analysis includes the detection and tracking of people and waves, as well as higher level analysis. We use the task of characterizing surfing conditions as a case study and demonstrate our ability to estimate the values of key parameters that can help determine the quality of a surf spot at a given time.**

***Index Terms*—beach webcams, people tracking, surfing video analysis, video monitoring, wave tracking**

I. Introduction

Water bodies, such as lakes, seas, and oceans, play a vital role ecologically, economically, and socially. For that reason, humans constantly interact with them. However, water bodies are a rapidly changing, unpredictable, and sometimes dangerous environments. One way for utilizing human use of water bodies while reducing the risks involved in human-water interactions is to develop technological tools for analyzing and predicting water body behavior. Such tools are important for maintaining human safety since they may identify potential hazards such as large waves, rip currents or submerged objects. Safety officials may use this information to warn swimmers or close sections of the beach. Such tools may also be used to spot distressed swimmers or monitor beach crowd size and behavior in order to identify potential safety risks such as overcrowding or rowdy behavior, which could lead to accidents or injuries. Automatic monitoring of beach video streams can also provide real-time observations of the beach environment, including water quality, weather conditions, and the behavior of wildlife and beachgoers, such as swimmers and surfers. This paper presents a novel approach for monitoring water bodies. Our solution utilizes existing beach webcams and applies computer vision algorithms based on deep neural networks to analyze the behavior of water bodies, humans, and the interactions between them.

Our approach for analyzing "beach scenes" combines the detection and tracking of people in a coastal environment with the analysis and tracking of waves. The first task of detecting and tracking people in a coastal environment was approached by researchers initially using classical image processing methods for object detection and tracking [1]. Recent studies on this task have employed deep neural networks for object detection [2]. For example, surfers were detected in videos obtained from beach webcams in [3] using saliency maps and probabilistic methods, and in [4] using Faster R-CNN, R-FCN and SSD deep neural networks. The second task of wave detection, tracking, and analysis is more difficult. Prior studies on this task have used videos from different sources for a variety of purposes, ranging from the general study of wave breaking dynamics [5]–[7] or coastal monitoring [8] for analyzing surfing conditions [9], [10]. This task, like the task of people detection, was first approached using classical image processing methods [5], [9], and later with deep neural networks [6]–[8], [10].

Beach webcams are typically used to provide live streaming video of the beach and surrounding area to give people a sense of the current conditions at the beach, such as the weather and the waves. Using deep neural networks for detection, and a tracking algorithm, we track both people and waves in videos acquired using existing beach webcams. Compared to previous work that only focused on one of these tasks, tracking both people and waves allows us to gain a better understanding of the ”beach scene”. We use the task of characterizing surfing conditions as a case study. To the best of our knowledge, this is the first work to detect, track, and analyze people and waves, and derive insights from analyzing them simultaneously. The proposed algorithm can distinguish between people in various poses and draw conclusions from this information. In our case study, we distinguish between a ”waiting surfer,” who sits or lies on his surfboard waiting for a wave, and a ”riding surfer,” who stands on his surfboard during a wave ride. These two unique abilities, tracking waves and people simultaneously, and distinguishing between a ”waiting surfer” and a ”riding surfer,” allow us to match a surfer with the wave he is riding. By doing so, we are able to gain unique insights about surfing conditions, using information such as the ratio of waves ridden to waves unridden, or the ratio of wave height to surfer height.

II. Proposed Solution

Fig 1 depicts a block diagram of our proposed solution for the surfing condition analysis task. For detection and

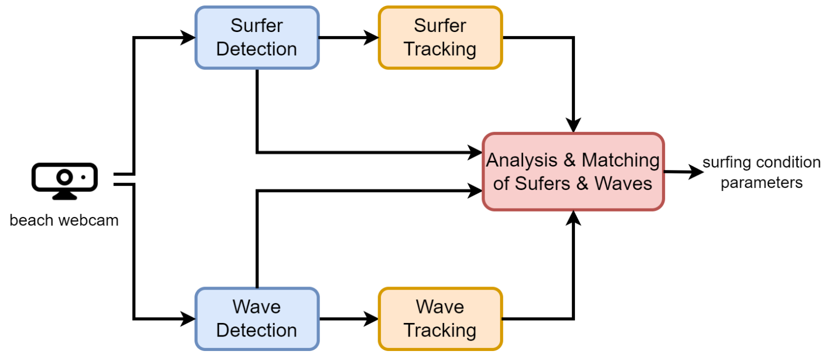


Fig. 1: Block diagram of our case study solution for surfing condition analysis.

classification, we use the Faster R-CNN [11] object detector with the ResNet [12] architecture as the backbone. The network outputs are bounding boxes of the detected objects as well as a label for each object. Faster R-CNN is a popular object detector that has shown impressive results in various computer vision tasks. It consists of two stages: a region proposal network (RPN) and a region-based convolutional neural network (R-CNN). The RPN is a fully convolutional network that generates a set of object proposals, which are regions of the image that potentially contain objects, while the R-CNN classifies and refines these proposals into actual object detections. One of the key advantages of Faster R-CNN over other object detectors is its accuracy. Additionally, it is fast and can process videos in real-time. With the use of the RPN, Faster R-CNN can efficiently generate a large number of object proposals, reducing the need for an exhaustive search over the entire image. The use of shared convolutional features between the RPN and R-CNN further reduces computation time and memory usage, resulting in faster and more efficient object detection.

We used a Faster R-CNN network pre-trained on the COCO [13] dataset and fine-tuned it by transfer learning with our dataset of surfing videos. Transfer learning is a popular deep learning technique that involves using a pre-trained model on a large dataset and fine-tuning it on a smaller dataset for a new task. This approach is particularly helpful when the dataset available for a given task is small, as is often the case in many computer vision applications, including ours. Through fine-tuning, where the pre-trained model is adjusted to suit the new task, a high accuracy can be attained with less training data and computational resources. The dataset we created for training contains 600 labeled frames from two beach webcams. It has, in total, 111 sitting surfers, 39 standing surfers and 36 waves.

For tracking we use SORT (Simple Online and Realtime Tracking) [14], a popular fast multi-object tracking algorithm. The algorithm consists of two main steps: detection and tracking. The detection step involves using an object detection algorithm, such as Faster R-CNN, to identify the objects of interest in the video frames. The tracking step involves assigning IDs to the detected objects and tracking them across frames to maintain their identity. SORT uses a Kalman filter to predict the position and velocity of each object in the video frames. The Kalman filter estimates the object's state (position, velocity, acceleration, etc.) based on the previous state and current measurement. SORT also employs the Hungarian algorithm to assign the detected objects to existing tracks. The Hungarian algorithm solves the data association problem by minimizing the cost of matching detections to existing tracks.

SORT is highly configurable, allowing fine-tuning of its hyperparameters to optimize performance. These hyperparameters include:

- *min_hits*: The minimum number of matching detections required to start a new track. Higher *min_hits* values make the tracker more selective, which can help eliminate false positives.
- *iou_threshold*: The minimum intersection-over-union (IoU) value required between a new detection and an existing track for them to be considered a match. Lower *iou_threshold* values make the tracker more permissive, which can help reduce the number of unmatched detections.
- *max_age*: The maximum number of frames a track can exist before terminating without a matching detection being found. Higher *max_age* values make tracking more persistent, which can help prevent track termination due to a temporary lack of detections, occlusions, or other disturbances.

In surfing scenarios, attaining continuous detection and tracking is challenging due to the water dynamics. However, surfers and waves in such scenarios are relatively sparse and often do not occlude each other. Therefore, we fine-tuned SORT hyperparameters accordingly. We increased the default values of *min_hits* and *max_age*, while decreasing the default value of *iou_threshold*. We also introduced a new hyperparameter that allows us to smooth out tracking even further as post-processing. Since we trained the neural network on a small dataset, detection is not always continuous, so the tracked wave or surfer is not detected in all video frames. As the appearance of surfers and waves in beach scenes is relatively sparse, matching an existing track with a new detection is possible even if there is no strict overlap between them. The new hyperparameter allows to enlarge the detection bounding boxes for each trajectory.

After the stage of analysis and matching of surfers and waves, the algorithm provides the following outputs:

- Number of waves ridden by surfers. Provides an estimate of the quality of the waves in terms of their suitability for surfing. The higher the number of ridden waves, the better are the surfing conditions.
- Number of waves unridden by surfers. This output is also relevant for estimating the quality of waves, especially compared to the number of waves ridden during the same period of time. A high number of unridden waves, compared to the number of ridden waves, may indicate poor wave quality due to factors such as low height, poor shape, or excessive ”chop.”
- Number of surfers: An estimate of the popularity of the surf spot. Can also be used to determine the surf spot's peak hours and how crowded it gets.
- Average ratio between the surfer height and the height of the wave he is riding. Useful for determining wave heights and can provide insights into the typical surfing classifications of wave heights, such as ”knee-high,” ”shoulder-high,” ”overhead,” or ”double overhead,” commonly used by surfers.
- Maximum ratio between the surfer height and the height of the wave he is riding. Can be used to determine the maximum surfable wave height for the surf spot.
- Average wave velocity. Provides information about the power and energy of the waves. The higher the wave velocity, the greater the energy of the waves, which can make for better surfing conditions.
- Average surfer velocity. Can be used to assess the skill level of the surfers. Experienced surfers tend to move faster and make more turns, whereas novice surfers move slower and with less control.

III. Results

Results were obtained from videos of beach scenes acquired from webcams not in the training set. Figure 2 depicts sample detection and tracking results for four beach scenes with diverse settings.

Table I presents analysis results for the four beach scenes from Fig. 2. This type of data can help surfers understand and compare the surfing conditions at multiple surf spots. For example, the Ericeira Beach appears to be an attractive surf spot as it has a relatively large number of surfers and a high ratio of ridden to unridden waves. On the other hand, the Rockaway Beach appears to be much less appealing for surfing as it has a relatively small number of surfers, and since no surfers riding waves were detected (therefore some of the data in the table for this beach are not available).

To quantify the accuracy of the outputs of the proposed algorithm, we manually annotated the same beach scenes but at different times than those at Table I. We present our ground truth annotations and the corresponding algorithm outputs in Table II. As can be seen in the table, the algorithm tends to overestimate wave detections. However, the results of the algorithm still allow a correct comparison between the quality of different surf spots.

TABLE I: Sample analysis results for the four beach scenes from Fig. 2. Surfer and wave speed is given in km/h. Average and maximum wave height is given relative to the height of the surfer riding the wave.

	surfers		waves				
surf spot	#	spd	ridden	unrid	avg ht	max ht	spd
Ericeira	35	17	9	2	1.4	1.6	15
Backdoor	30	25	8	4	2.4	3.5	22
El Porto	10	18	5	4	1.3	2.0	17
Rockaway	6	N/A	0	9	N/A	N/A	20

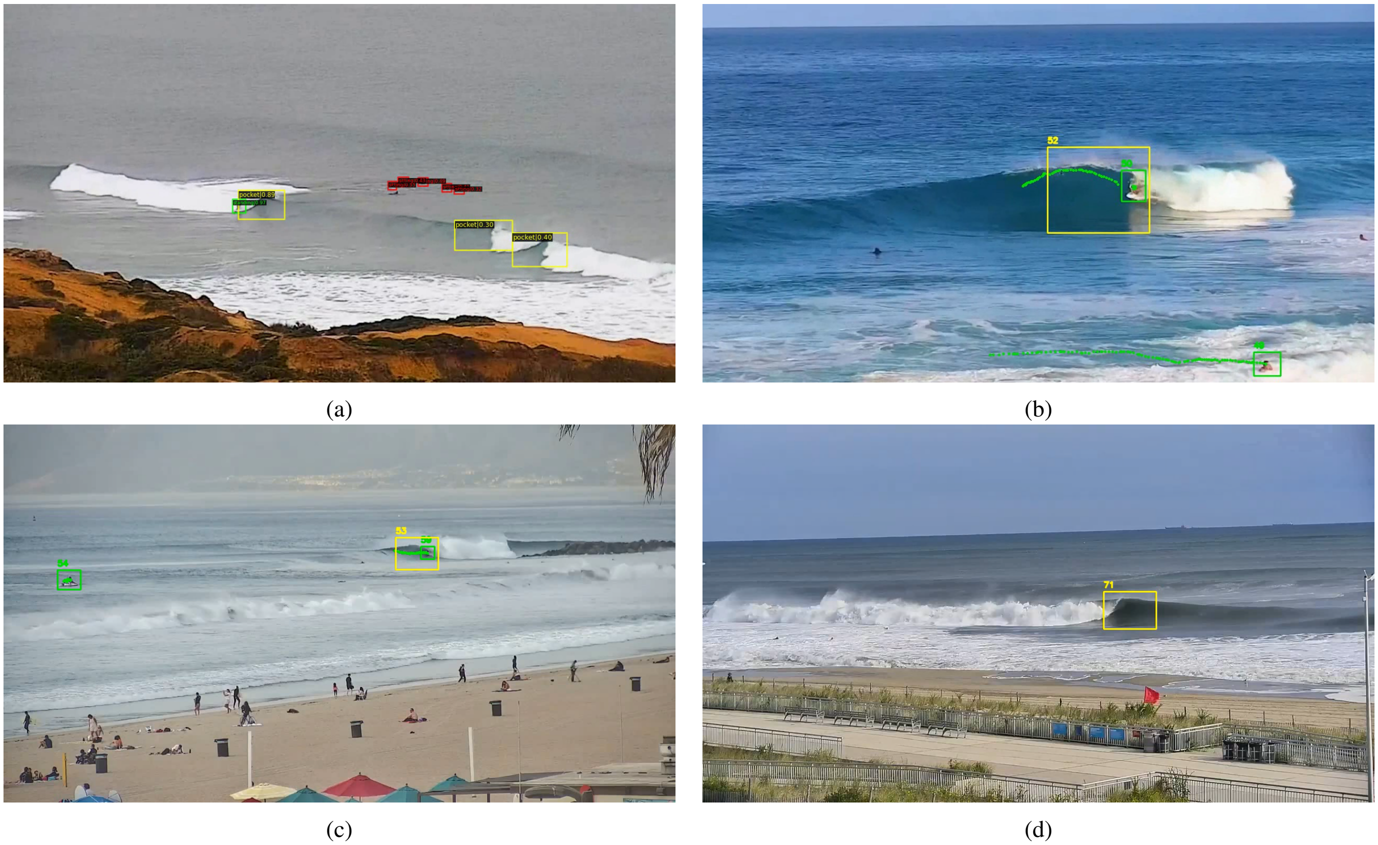
TABLE II: Proposed algorithm outputs vs. ground truth values for the number of waves ridden and the number of waves unridden by surfers. The values are for the same beach scenes at Table I but at different times.

	ground truth		algorithm outputs	
surf spot	ridden	unridden	ridden	unridden
Ericeira	9	3	17	3
Backdoor	12	0	13	3
El Porto	4	3	5	4
Rockaway	3	6	0	9

The reason behind the algorithm's overestimation of the number of ridden waves can be attributed to the tracker creating new tracks for the same surfer during a single wave ride, which results in counting a single ride as several. This issue can be mitigated by improving the detection accuracy through the use of a larger training dataset, and by switching to a more advanced tracker. Similarly, the algorithm's overestimation of the number of unridden waves can be attributed to its tendency to mistakenly recognize different ”pockets” of a single wave as separate waves. This issue can be mitigated by employing a more aggressive post-processing method that eliminates short tracks of temporal ”pockets” of the same wave. Rockaway Beach has different settings from the beach scenes in the training set. Despite this, the algorithm detected sitting surfers and waves on this beach with high accuracy. However, it encountered difficulty in detecting surfers riding waves, leading to all waves being classified as unridden. This issue can also be resolved by using a larger training dataset.

IV. Conclusions

This paper presents a novel approach for automatic monitoring of beach video streams using deep neural networks to analyze human activity and water behavior simultaneously. By integrating these two types of analysis, our approach provides new insights that were previously unattainable when analyzing each separately. Specifically, we demonstrated our approach by characterizing surfing conditions as a case study and showed that it can estimate key parameters that determine the quality of a surf spot, such as wave count and the ratio



(a) (b) (c) (d)

Fig. 2: Examples of detection (a) and tracking (b)-(d) in surfing scenes from (a) Ericeira, Portugal in October 2021 (b) Backdoor Reef Brake, Oahu, Hawaii in December 2021 (c) El Porto beach, California in January 2021 (d) Rockaway Beach, New York in October 2019. A yellow bounding box indicates a wave "pocket", a red bounding box indicates a "waiting surfer", and a green bounding box indicates a "riding surfer". A green trails indicates the track of a riding surfer in the last frames.

of ridden to unridden waves. Our approach has the potential to improve safety, environmental monitoring, research, and education related to beach activities using existing webcams already installed at many beaches.

Acknowledgment

The authors thank Prof. David Malah and the entire staff of the Signal and Image Processing Laboratory (SIPL) at the Technion for their ongoing assistance and support.

References

[1] R. L. da Silva, S. Chevtchenko, A. A. de Moura, F. R. Cordeiro, and V. Macario, "Detecting people from beach images," in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2017, pp. 636–643.

[2] S. Chevtchenko, R. Vale, F. Cordeiro, and V. Macario, "Deep learning for people detection on beach images," in *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2018, pp. 218–223.

[3] R. Bastos, "Monitoring of human sport activities at sea," Ph.D. dissertation, Master Thesis, Portugal, 2015.

[4] J. Fung. (2022) Surfer detection using a convolutional neural network (CNN) in Google's Tensorflow. [Online]. Available: https://github.com/citrusvanilla/surferdetection

[5] P. Green, "Investigation into the use of surf web cameras to improve the accuracy of forecasting wave behavior," *University of Edinburgh, Edinburgh*, 2018.

[6] J. Kim, J. Kim, T. Kim, D. Huh, and S. Caires, "Wave-tracking in the surf zone using coastal video imagery with deep neural networks," *Atmosphere*, vol. 11, no. 3, p. 304, 2020.

[7] C. Eadi Stringari, P. Veras Guimarães, J.-F. Filipot, F. Leckler, and R. Duarte, "Deep neural networks for active wave breaking classification," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

[8] G. Scardino, G. Scicchitano, M. Chirivì, P. J. Costa, A. Luparelli, and G. Mastronuzzi, "Convolutional neural network and optical flow for the assessment of wave and tide parameters from video analysis (leucotea): An innovative tool for coastal monitoring," *Remote Sensing*, vol. 14, no. 13, p. 2994, 2022.

[9] S. Mckenzie and R. Green, "Gnarometer surfcam live inspector," in *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2018, pp. 1–6.

[10] M. Thompson, I. Zelich, E. Watterson, and T. E. Baldock, "Wave peel tracking: A new approach for assessing surf amenity and analysis of breaking waves," *Remote Sensing*, vol. 13, no. 17, p. 3372, 2021.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[14] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016, pp. 3464–3468.