Visual analysis of video recorded dinghy sailing sessions

Gijs M.W. Reichert* Marcos Pieras* Ricardo Marroquim* Anna Vilanova †

* Delft University of Technology, Delft, The Netherlands.
† Eindhoven University of Technology, Eindhoven, The Netherlands

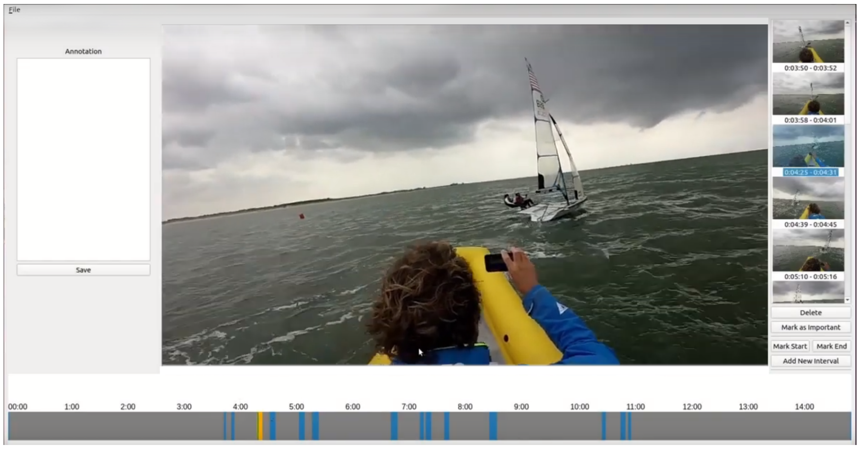


Figure 1: The Visual Analysis Framework. The timeline (bottom), list of intervals and tools (right) support the coaches in reviewing and annotating (left) recorded training sessions.

Abstract

Nowadays video plays an important role in the coaching of athletes across many different sports. For sailing training sessions, the videos are recorded from the coach boat and provide ways to review and analyze the training sessions aiming at improving the sailors performance. On one hand, videos are commonly recorded with handheld devices that are rather cumbersome to acquire and are prone to missing many important moments. On the other hand, recordings of entire sessions with an integrated camera in the coach boat are difficult to analyze given their length and the lack of image stability. We present a pipeline to facilitate the visual exploration of these full session videos. We extract manoeuvres as interesting points from the recordings, and provide a visualization framework to present the video and processed data. Manoeuvres are extracted by detecting and tracking the boat and sailors. With the visualization tool, the user can locate and visually inspect those manoeuvres for coaching tasks. We evaluated the potential of the framework and from the results we conclude that the manoeuvre detection is reasonably accurate and some coaches see potential in the presented framework.

1 Introduction

Nowadays technology and data analysis are becoming increasingly intertwined with sports. The data can help assess performance during training and in competitive settings.

This trend holds for the complex sport of sailing as well, where more and more sensors are added to the boats to measure the performance. However, for the Olympic dinghy class the use of sensors during training is not standard practice, because there is no clear strategy to analyze the data and the sensors can limit the ability to move freely in the boat. Moreover, the use of sensors is not allowed during races. To record and review their performance the coaches and athletes rely mostly on videos. Training sessions usually consist of $2-3$ hours where the coach follows the sailors around in a rigid inflatable boat (RIB) and assesses their handling and technique.

Manoeuvres are an integral and paramount part of sailing. Here, by manoeuvres we refer to tacking and jibing. Tacking is when a boat turns the bow of the boat in the direction the wind is coming from and then keeps turning “through the wind” to catch wind on the other side of the sail. Jibing is the opposite of tacking. During these manoeuvres the sailors usually switch from one side of the boat to the other as the sail also switches sides. During races every second counts, and whenever tacks and jibes are not performed perfectly a sailor can lose precious distance to the competitors. Both the decision when to make a manoeuvre and its execution are important in being faster than your opponents, hence, coaches make short video clips of manoeuvres to further discuss with the sailors. They use handheld recording devices, such as smartphones, and usually each clip is not longer than a minute. Nevertheless, since the videos are recorded in a rather cumbersome manner using handheld devices, many manoeuvres are missed as the coach does not have enough time to capture all of them. Acquiring the complete set of manoeuvres performed during the training sessions would allow for better analysis of the sailor's performance.

In order to avoid the burden of using handheld devices and to allow for capturing the complete set of manoeuvres, the coaches are currently switching to using mounted cameras on the RIB that record the entire training session. Although this new setup can potentially capture all manoeuvres, it presents a new issue as going through the entire recording to find the manoeuvres after every training is undesirable and too time-consuming for the coaches and sailors.

The main contribution of this work can be summarized as follows: A pipeline to provide visual analysis of the recorded videos. It includes the extraction of manoeuvres from recorded footage, highlighting potentially interesting segments in time, which can in turn be explored and annotated using the created visual interface.

2 Related Work

The aim of video visualization is not intended to provide fully automatic solutions to the problem of making decisions about the contents of a video. Instead, it aims at offering a tool to assist users in their intelligent reasoning while removing the burden of view-

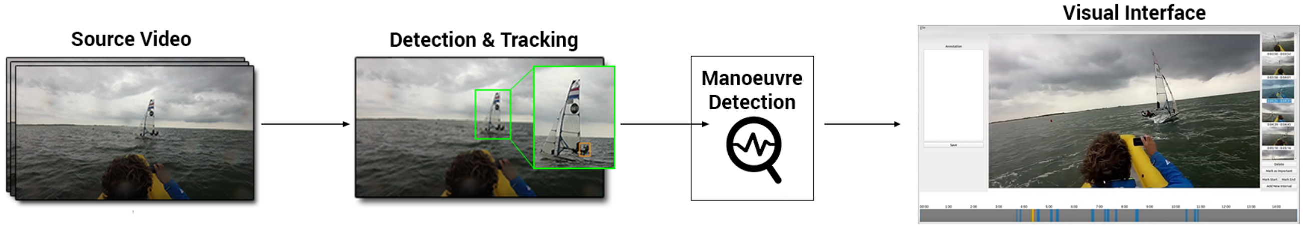


Figure 2: The implemented sailing manoeuvres analysis pipeline to locate manoeuvres in a video and visualize these intervals.

ing videos [3]. This reasoning can be translated to several visual task [15]: video annotation, browsing, editing, navigation, recommendation, retrieval, and summarization. Our approach combines features from annotation, navigation, and summarization. The main objective is the exploration of interesting events on video, the location of *keyframes*. This location can be carried out in a automatic or manual fashion.

In [7], authors developed a visual strategy to analyze semantically and cinematographically movies with the usage of video and text data. Similar, in [4] a visual analysis tool is design to analyze long sequences with multiple events location on a challenging data as it is in first person video.

Numerous sports vision-based analysis approaches exist, as can be seen in the survey by Barris et al. [1]. For example, to provide feedback to athletes, as described in the work of Donoghue et al. [12]. This holds for sailing as well, where video is used to review the performance and technique and provide feedback to the sailor.

The usage of video in sports visual analytics has two roles, the passive one, where it complements other types of data. For instance, Plok et. al. [13] use video to reinforce learning outcomes after analyze tabular data, as one coach said during their evaluation, *seeing is believing*. On the other hand, the usage of video as an active role, where it is the main data input, like in [9], authors developed a visual analytics tool for multiple *keyframes* annotation with glyph techniques.

3 Sailing Manoeuvres Analysis Pipeline

To be able to find, extract and visualize manoeuvres we need to go through a number of steps that we refer to as the Sailing Manoeuvres Analysis pipeline. An overview of the pipeline is illustrated in Figure 2. As mentioned before, the sailors switch sides during a manoeuvre. We hypothesize that this is the most promising cue to detect a manoeuvre from video. We assume the coach is following the sailing boat from behind with the recording camera facing forward. We track the location of the boat and sailors (i.e., Detection & Tracking in Figure 2), and subsequently use this information to detect the manoeuvres. The last step of the pipeline is visualizing the detected manoeuvres. Our interface allows the coaches and sailors to efficiently navigate through a recorded training session, analyzing and annotating the most interesting manoeuvres.

3.1 Detecting and Tracking the Boat and Sailors

To detect manoeuvres we use the location of the boat and the location of the sailors with respect to the boat. For the detection of sailors and boat, we use a pre-trained neural network for object detection that contains both classes of interest. For a review on this topic please refer to Zhao et al. [17].

We selected MobileNet because it is an efficient model, originally designed by Google as a light weight deep neural network that could be used on mobile devices [5] [6], and a pre-trained model is readily available [1]. The pre-training was executed on the Microsoft Common Objects in Context (COCO) dataset [10], which contains images for the classes *Person* and *Boat*.

By sequentially feeding the frames of a video to the network we obtain the bounding boxes around our regions of interest (ROI) for each frame, for our application the boat and persons. However, during experiments we realized that our objects of interest are often not detected in the frames. This is especially true for the sailors. To avoid frames where the boat or sailors are not detected we combine the detection with an extra tracking algorithm. The tracking method used was the Discriminative Correlation Filter with Channel and Spatial Reliability (DCR-CSF) [11]. The part of the frame inside the bounding box output of the network is used to initialize the filter of the DCR-CSF tracker. Once initialized, we use the tracker to follow the boat and sailors. The tracker is still not perfect and prone to drift over time due to accumulated tracking errors. To compensate for this drift, we compare the distances between the centers of the bounding boxes from the network and the tracker, and we re-initialize when the distance is larger than a given amount of pixels. In our experiments we used a 25 pixels threshold for a 1280x720 footage. However, if the network also fails to detect the boat or sailors we assume that they are either too far away or outside the frame, and the tracker cannot be initialized. If the tracking filter was already initialized using the detected object of interest the tracker will continue tracking. During experiments we disabled the tracker whenever the network did not detect a boat or person for more than 60 sequential frames, which empirically worked well and helped to avoid unreliable location data.

The quality of the videos regarding stability is low, since the cameras are attached to the RIB which is not a stable platform. We added a stabilization step in order to improve the tracking accuracy. The videos were stabilized based on the horizon using a method similar to the Horizon-tracking method of Schwendeman et al. [16]. We noted that in our experiments the accuracy of detection and tracking increased on average by $10-15\%$ using the stabilized video. Additionally, we hypothesize that a stabilized video is easier to analyze visually by a human.

3.2 Manoeuvre Detection

The goal of the Manoeuvre Detection stage of the pipeline is to detect clips when sailors switch sides. It utilizes the detected labeled bounding boxes surrounding the objects of interest. The centre of the boat's and sailor's bounding boxes are used to calculate the signed distance between the sailors and the centre of the boat, d. We only consider the horizontal difference (x-axis) which indicates the switching of sides. Even if the boat is not straight up the sailors should still be far from the middle of the boat, especially because the sailors will be Hiking (hanging out of the boat) to keep the boat as straight up as possible.

The calculated d is noisy but the crossing moment is still noticeable, as can be seen in Figure 3 where we visualize the signed distance of the sailors to the center of the boat, d, through time. Since we assume the video is registered from behind the boat, we define the vertical line in the middle of the bounding box as the mid-

[1] https://github.com/chuanqi305/MobileNet-SSD/

	Question	Answer Type
Q1	Which video (1 or A) is easier to analyze? (1) - Strong preference for Video 1 ; (3) - Both videos are equally easy/difficult; (5) - Strong preference for Video A	Option 1-5
Q2	What aspects of this framework, if any, would be useful for coaching?	Open Question
Q3	What features are you missing in this framework?	Open Question
Q4	What features are not useful?	Open Question
Q5	How likely is it that you would use this in coaching?	Likert Scale
Q6	How useful, in your opinion, is the timeline with marked intervals/manoeuvres in the framework?	1-10 Scale

Table 1: Questions asked to the coaches as part of the User Study.

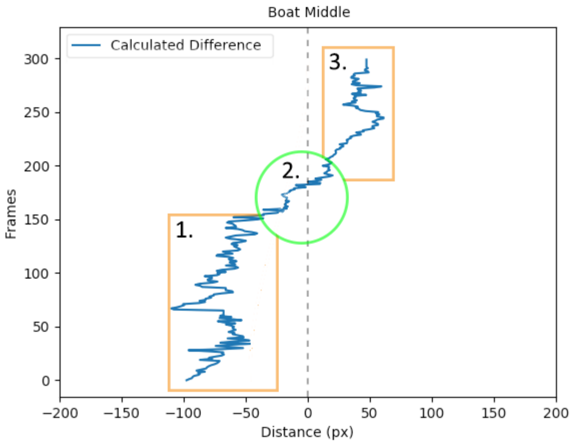


Figure 3: Three parts of the manoeuvre. (1) Stable on left side, (2) Crossing the middle of the boat during manoeuvre and (3) Stable on the right side.

dle of the boat. Whenever the sailors cross this line, and remain on the other side, we assume that a manoeuvre has just occurred. The three important moments of the manoeuvre detection can be seen in Figure 3, and the entire duration of the manoeuvre is bounded using the "stable" locations. In other words, we search for the point where the sailors start to move to the other side of the boat and the point where the sailors remain on the other side of the boat.

An option to remove the noise from the data and detect the stable locations and crossing, would be a sliding window, for which a size and a fixed smoothing parameter σ need to be defined. When using a high fixed value for sigma, for example $\sigma = 140$ as in Figure 4, we get a smooth curve but the zero-crossing point has an offset of more than 50 frames with the real crossing. To avoid the definition of a specific *sigma* value, we adapted a scale space technique called Edge Focusing [2] to filter the data and locate the zero crossings. With our adaptation of Edge Focusing, we are able to robustly pinpoint the frame where the crossing occurs. More specifically, we compute a Laplacian of Gaussian (LoG) filter with σ in the range $[e^a, e^b], a = 5, b = 0$ and a stepsize of 0.005, as suggested by Haar Romeny [14]. The output of the LoG is calculated for every value in this range and the frame numbers of the zero-crossings are stored as the "signatures", as illustrated in Figure 5. The small stepsize ensures the difference in frame number in the positive and negative edge between neighboring signatures is never larger than 1 frame in time. Then, the positive edge is tracked from coarse to fine scale to find the frame where the sailors cross over to the other side (see orange arrows in Figure 5).

To determine if the sailors remain at a stable position on one side of the boat we fit a Least Squares Regression Line [8] to the output of the Detection & Tracking stage. More specifically, we look at the slope of the line that we fit to the horizontal change of the sailors for a window of the average time it takes to perform half a manoeuvre.

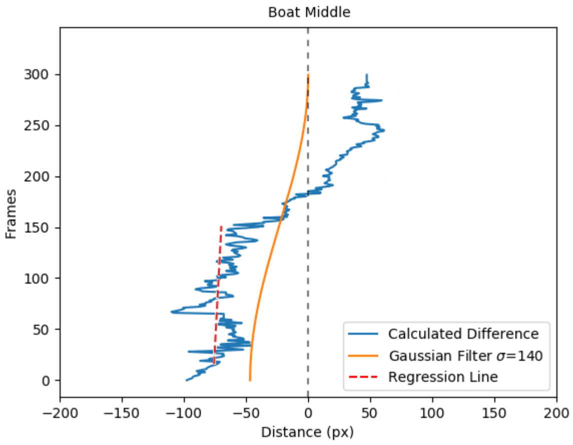


Figure 4: Regression line fitted to noisy data to find stable point marking start of manoeuvre, before zero crossing (middle of manoeuvre). Orange line is the output of the (too) coarsely filtered Calculated Difference.

For 30Hz videos this implies in around 90 sequential frames. When the fitted line has a slope of less than 0.15 radian, we judge the location of the sailors to be stable with respect to the middle of the boat. In other words, fitting the line over a large enough window reduces the influence of the noisy data and when the slope of the line increases we assume the sailors are moving to the other side. The slope of the line should then decrease again when the sailors position stabilizes again on the other side. The last stable frame before the crossing, and the first stable frame after the crossing mark the manoeuvre interval.

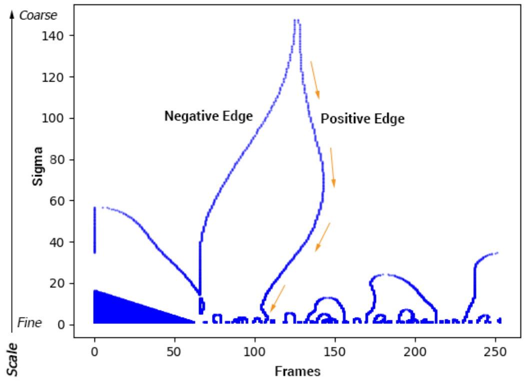


Figure 5: Adaptation to frames of Edge Focusing "signature" graph at different values for σ, allowing tracking from coarse to fine.

3.3 Visual Analysis Interface

For the exploration of a recorded session with the detected manoeuvres, we provide a visual analysis interface. We designed an interface which displays the video and the time intervals that have been detected as interesting, as depicted in Figure 1. We show the video of a recorded session, where the user has the option to watch the original or stabilized version of the video. The time intervals are shown in a timeline, where we use contrasting colors to highlight the location in time of the intervals (Blue) and the currently selected interval (Orange). Next to this, the user also has the ability to select the intervals from a list of thumbnails, where each thumbnail is labeled with the timestamps marking the beginning and end of the interval. The list of thumbnails and the timeline are linked, clicking an interval in the timeline will highlight (Orange) the interval in the timeline and at the same time highlight the thumbnail in the list, and vice versa. The interface contains tools to perform a few different operations that are useful in the debrief after a training session. The user has the ability to add and delete intervals, mark intervals as important and annotate selected intervals using the text-box on the left-hand side of the interface.

The timeline with contrasting colors gives a good overview of the potentially interesting events during the session. Next to this, the thumbnails give an indication of what is happening in the time-interval. This makes it easier to select an interval that is interesting and remove false positives created by the manoeuvre detection stage of the pipeline.

With this interface, the user can explore the manoeuvres on the video, specifically, the location during the session and its duration. With the provided functionality, the user can visual inspect and compare different manoeuvre and reinforce the coaching process.

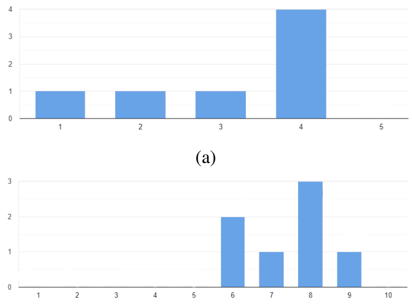
4 Evaluation

The manoeuvre detection pipeline was evaluated using three representative test videos and a set of 27 recorded manoeuvres. We removed the situations where our assumptions do not hold, e.g. RIB was not following the sailing boat from behind. Against this ground truth data of manoeuvres, the average sensitivity of the method is 72.72%. The results give an indication of what can be expected, when adhering to the assumptions of the method. When the assumptions do not hold false positives increase considerably, from 4 false positives to 22 in this case. False positives need to be discarded then manually in the visual interface.

For 20 out of a set of 27 manoeuvres, constructed using multiple videos of different lengths, the Adapted Edge Focusing method was accurate to the exact frame. The complete set of 27 had a median offset of 26 frames and an average of 45 frames. In practice, with a frame-rate of 30 Hz, we would be off by about a second which is negligible. The amount of data available is not large enough to draw strong conclusions, but does give an indication of the potential of the manoeuvres detection to be included in the visual analysis system for coaching.

To evaluate the developed framework a user study was also conducted with seven coaches. First, to evaluate if there is a preference for either stabilized or the original video, the users were shown three pairs of videos, each pair consisting of the original and stabilized versions. Then, for each pair the user was asked question **Q1** from Table 1. 10 of the 21 votes either prefer or strongly prefer the stabilized version of the videos. Motivations given for these choices in favor of the stabilized versions is that it makes looking at the details easier and that “the movements of the video are caused by the RIB, which are totally irrelevant”. We observe that the stabilized version is considered to be the better version by some coaches, but in future work the distracting moving border edges should be addressed.

Three of the coaches mentioned that all of the aspects of the framework could be useful in response to **Q2** in Table 1. Others named specifically the ability to mark manoeuvres and making notes that are attached to these intervals.



(a)

(b)

Figure 6: (a) Responses to Q5 from Table 1 on Likert scale. (b) Responses to Q6 from Table 1 on a scale from 1 to 10.

Two of the coaches, stated for **Q3** from Table 1 that they were missing an easy way to share and store the clips. Another feature that was mentioned was the ability to draw on the videos, as well as labeling/naming the clips. The prototype did not contain such functionalities, as they fall outside the scope of this work, but they could be incorporated at a future stage. One coach mentioned the ability to zoom in on the boat to be able to see more detail, because for the relatively fast sailing boats it is difficult to stay close during training.

Although we assumed that the thumbnails would give an indication of the contents of the interval, this is not the case if we consider **Q4** from Table 1. One coach stated that everything looks the same in the thumbnails and therefore it is not useful. This in line with the comments of another coach on **Q3**, who stated that you need a way to label/name the thumbnails to be able to distinguish between them.

The results for **Q5** and **Q6** can be found in Figure 6a. On average we can conclude that the coaches would likely use the framework in coaching, as 4 coaches voted 4 on a scale of 5. The main issues of the coaches that score low are matters outside the scope of this work, most of these issues were related to distance between camera and target boat or video quality leading to problems with observing details. With an average of 7.43 in **Q6** the coaches think the timeline is a useful addition.

5 Conclusion and Future Work

We presented a pipeline to improve the visual analysis of manoeuvres in recorded training sessions. Manoeuvres are automatically detected and visualized in the Visual Analysis Framework. It allows to efficiently locate manoeuvres in the recorded videos, without having to search through the entire session manually. This improves the analysis of the training sessions using only the video data. The user study indicates that this initial framework is promising and could contribute to the analysis of sailing training sessions.

Nevertheless, the manoeuvre detection would benefit from a more accurate detection and tracking of boat and sailors. Moreover, combining the video data with more data such as boat speed captured by sensors could further improve the Visual Analysis Framework. Next to this, further evaluation is needed by testing the pipeline on more videos and evaluating the experience of more coaches.

References

[1] S. Barris and C. Button. A review of vision-based motion analysis in sport. *Sports Medicine*, 38(12):1025–1043, Dec 2008. doi: 10.2165/00007256-200838120-00006

[2] F. Bergholm. Edge focusing. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):726–741, 1987.

[3] R. Borgo, M. Chen, B. Daubney, E. Grundy, G. Heidemann, B. Höferlin, M. Höferlin, H. Leitte, D. Weiskopf, and X. Xie. State of

the art report on video-based graphics and video visualization. *Computer Graphics Forum*, 31, 12 2012. doi: 10.1111/j.1467-8659.2012.03158.x

[4] K. Higuchi, R. Yonetani, and Y. Sato. Egoscanning: Quickly scanning first-person videos with egocentric elastic timelines. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6536–6546, 2017.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[6] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7310–7311, 2017.

[7] K. Kurzhals, M. John, F. Heimerl, P. Kuznecov, and D. Weiskopf. Visual movie analytics. *IEEE Transactions on Multimedia*, 18(11):2149–2160, 2016.

[8] M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li, et al. *Applied linear statistical models*, vol. 5. McGraw-Hill Irwin New York, 2005.

[9] P. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. *Computer Graphics Forum*, 31(3):1255–1264, 2012. doi: 10.1111/j.1467-8659.2012.03118.x

[10] 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*, pp. 740–755. Springer, 2014.

[11] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan. Discriminative correlation filter with channel and spatial reliability. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[12] P. O'Donoghue. The use of feedback videos in sport. *International Journal of Performance Analysis in Sport*, 6:1–14, 11 2006. doi: 10.1080/24748668.2006.11868368

[13] T. Polk, D. Jäckle, J. Häussler, and J. Yang. Courttime: Generating actionable insights into tennis matches using visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26:397–406, 2020.

[14] B. M. H. Romeny. *Front-end vision and multi-scale image analysis: multi-scale computer vision theory and applications, written in mathematica*, vol. 27. Springer Science & Business Media, 2008.

[15] K. Schoeffmann, M. A. Hudelist, and J. Huber. Video interaction tools: A survey of recent work. *ACM Comput. Surv.*, 48(1), Sept. 2015. doi: 10.1145/2808796

[16] M. Schwendeman and J. Thomson. A horizon-tracking method for shipboard video stabilization and rectification. *Journal of Atmospheric and Oceanic Technology*, 32(1):164–176, 2015.

[17] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 2019.